January 2019

Structural Insights Into Histone Modifying Enzymes

Shruti Amle
Wayne State University, amle.shruti@gmail.com

STRUCTURAL INSIGHTS INTO HISTONE MODIFYING ENZYMES

by

SHRUTI AMLE

THESIS

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

2019

MAJOR: BIOCHEMISTRY AND MOLECULAR BIOLOGY

Approved By:

_____

Advisor Date

_____

_____

_____

ACKNOWLEDGEMENTS

Writing this thesis has been extremely captivating and gratifying. I take this opportunity to express my deep and sincere acknowledgements to the number of people for extending their generous support and unstinted help during my entire study.

Firstly, I would like to express my respectful regards and deep sense of gratitude to my advisor, Dr. Zhe Yang. I am extremely honored to study and work under his guidance. His vision, ideals, timely motivation and immense knowledge had a deep influence on my entire journey of this career. Without his understanding and support, it would not have been possible to complete this research successfully. I also owe my special thanks to my committee members: Dr. David Evans and Dr. Ladislau Kovari for their valuable insights and suggestions that have been very helpful throughout my project.

I am highly thankful to my former lab members Nicholas Spellmon and Joshua Holcomb for their valuable support and help to learn the research techniques in the lab. Their advices, criticism and guidance have added considerably to my experience. I extend my thanks to my other lab friends Yingxue Zhang, Maysaa Doughan, and Emerson Perry for outstanding co-operation and sharing their knowledge with me. Apart from that, I am also thankful to the members outside my lab: Kendall Muzzarelli for teaching me different lab skills and Chandni Patel for helping me find out specific reagents and proper instruments for my experiments. I am grateful to all of those with whom I have had the gratification to work during this and other related projects.

I am especially obliged to April Wolak who is a student records clerk at Biochemistry and Molecular biology department. She has been supportive of providing academic assistance and paperwork I needed throughout my master's studies and was extremely helpful in placing lab orders.

Last but not the least, I feel the deep sense of gratitude towards my family for their love and sacrifices for educating and preparing me for the future. I wish to thank my husband for his support and unending inspiration in the pursuit of this project.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: ROLES OF CHROMATIN-MODIFYING ENZYMES IN GENE REGULATION

1.1 General Introduction

1.1.1 Epigenetic Regulation

Nearly all cells of a multicellular organisms comprise the same DNA sequence and are genetically identical but are structurally and functionally diverse due to the variations observed in gene expression [1]. Besides a person's DNA sequence, there is an additional level of transcriptional control that decides which genes remain turned on and which remains turned off [2]. Epigenetic regulation involves the alteration of DNA and the proteins that are associated with DNA. This results in changes to the conformation of DNA and accessibility of other factors to DNA, without a change in actual DNA Sequence [2, 3]. Epigenetic regulation centers on chemical modifications that take place on DNA or specific amino acids in histone proteins that DNA is wrapped around. Three common forms of epigenetic regulation are 1) DNA methylation, 2) histone post- translational modifications and 3) RNA based mechanisms (micro-RNAs) (figure 1.1) [3-5].

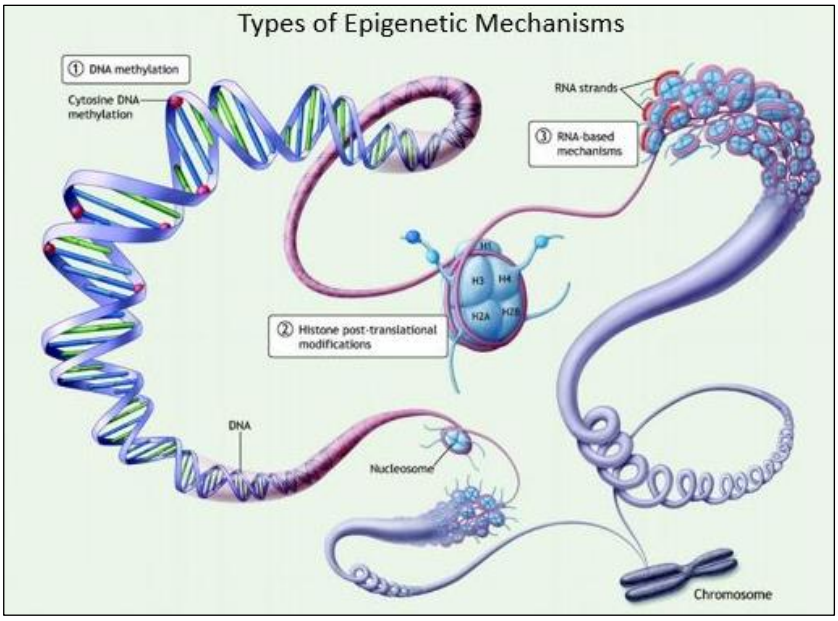


Figure 1.1: Summary of the three primary epigenetic mechanisms [6]

(1) DNA methylation (2) Histone post translational modifications. (3) RNA-based mechanisms.

DNA methylation is one of the epigenetic mechanisms used to regulate the gene expression which includes covalent transfer of methyl group catalyzed by DNA methyltransferases (DNMTs) to the cytosine ring of DNA molecule on C-5 position [7-10]. DNA methylation takes place in different biological processes, for instance, inactivation of X-chromosome, embryonic development, suppression of genes, chromosome stability, and carcinogenesis and plays an important part in maintenance of DNA stability and integrity [8, 11].

1.1.2 Histone Post Translational Modification:

DNA is wrapped around eight histone proteins into the structures known as nucleosomes. Each nucleosome possesses copies of four core histones (H2A, H2B, H3, and H4) [12]. Histone tails undergo different post translational modifications catalyzed by different enzymes that recognize specific target sequences and alter the interaction of histones with DNA and nuclear proteins. This leads to epigenetic modifications for regulating several normal and

disease-related processes. Histone post-translational modification has its impact on gene expression and can also play an important role in diverse biological processes such as transcriptional activation and inactivation, chromosome packaging, and DNA damage/repair [12]. Methylation, phosphorylation, acetylation, ubiquitylation are the most common covalent post-translational modification (PTM) to histone proteins [13-15].

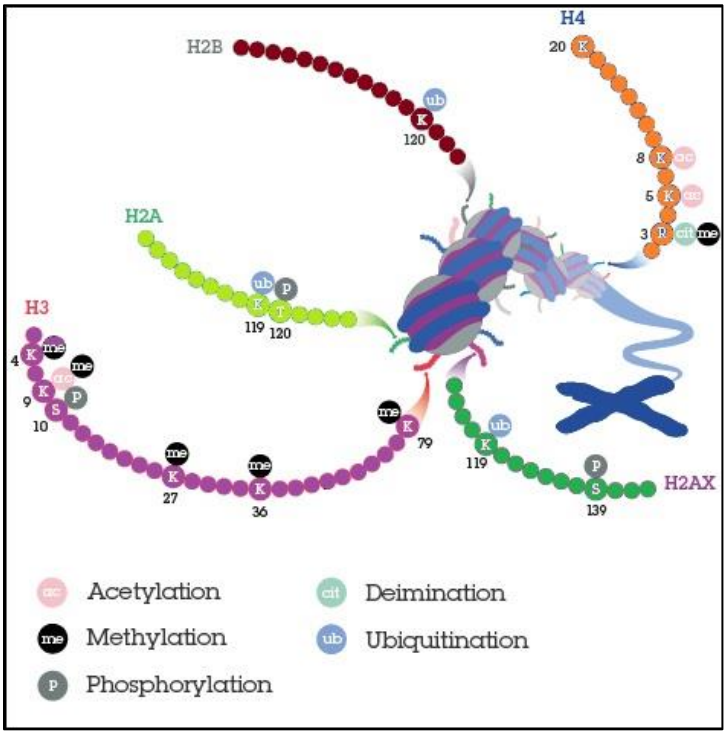


Figure 1.2: Most common post translational modifications to histone proteins [16]

1.1.3 Histone Methylation

It is the process that involves addition of methyl groups to the histone proteins mediated by histone methyltransferases (HMTs). This epigenetic modification mainly occurs on histone tails to the lysine and arginine residues which leads to gene activation or repression based on the amino acid residue that is methylated. Arginine residues may be mono- or di- methylated whereas, lysine residues may be methylated once, twice or thrice. There are several different histone methyltransferases known to methylate specific arginine or lysine residues. Depending upon where

the methyl groups are transferred by the HMTs, the name changes. So, if the methyl group is transferred to the lysine residue they are known as lysine methyl transferases (HKMT) and if it is transferred to an arginine residue they are known as arginine methyl transferases (HRMT) [17, 18].

1.1.4 Histone Ubiquitylation

It is an enzymatic post-translational modification that involves addition of a ubiquitin molecule to a substrate protein most commonly to the lysine residue of a substrate [19]. Ubiquitin is an 8-kDa polypeptide found in eukaryotic cells. It is present either in free state or is found covalently linked to other proteins [20]. This is a conserved eukaryotic protein that is attached to histone lysine via successive action of three enzymes, E1- UB activating enzyme, E2- UB conjugating enzyme and E3- UB ligase [20, 21]. Both, the substrate specificity (i.e., where lysine is targeted) and the degree of ubiquitylation (i.e., either mono- or poly-ubiquitylated) is regulated by the enzyme complexes. Histone H2A and H2B are two most highly ubiquitylated proteins found in the nucleus. Histone ubiquitylation plays a central role in the DNA damage response and in many other cellular processes, including transcriptional regulation, protein degradation along with protein-protein interactions. My research focus on the structure and function of ubiquitin-conjugating enzyme as well as lysine methyltransferases in human and yeast.

1.1.5 Histone Lysine Methyltransferases:

Histone lysine methyltransferases (HKMT) is a part of Histone methyltransferases (HMT) that catalyze the transfer of methyl groups from the co-factor SAM (*S*-adenosylmethionine, AdoMet). AdoMet comprises of methylthiol group which is highly reactive. This methyl group is then added to the ε-amino group of a lysine residues on histones H3 and H4 [22, 23] forming mono-, di-, or tri- methylated products in addition to its analogue adenosylhomocysteine

(AdoHcy). Once the methyl marks are added to the lysine residue, it can result in regulation of protein function by either recruiting binding proteins or direct effects. [24]. Lysine methyltransferases (KMT) can be classified into SET domain-containing and non-SET domain-containing proteins. SET domain is a highly conserved catalytic unit and found the in majority of known KMTs. Several non-SET-domain KMTs have also been identified e.g. DOT1L, METTL10 etc. SET-domain containing proteins have become more important in epigenetics and have been linked to cancer. They are involved in numerous diseases and play a crucial role in regulating gene transcription. Although, the functions of histone methyltransferases have been explored, the pathological mechanisms and downstream of lysine methylation remain to be elucidated [24].

1.1.6 SMYD Proteins

SMYD protein family is an important class of lysine methyltransferases which catalyzes methylation reaction on both histones and non-histone targets [25]. At present, five members of SMYD family including SMYD1, SMYD2, SMYD3, SMYD4 and SMYD5 have been identified which plays diverse biological functions and are involved in epigenetic transcription regulation, cancer proliferation and the immune system [26]. Two highly conserved structural and functional domains, the SET and MYND domains is present in all the members of the SMYD protein family. This conserved domain plays a role in lysine methylation and protein-protein interaction, respectively. SET (Suppressor of variegation, Enhancer of Zeste, Trithorax) domain in N-lysine methyltransferases is approximately 130 aa long and is a highly conserved catalytic motif responsible for methyltransferase activity by adding methyl marks to lysine residues of proteins using SAM as a methyl group donor [27]. SET domain is “split” by MYND (Myeloid-Nervy-DEAF1) domain. It is a cysteine-rich zinc finger motif and is mainly involved in protein-protein interaction with preference for binding with a proline rich motif (PXLXP). It is well-defined by a

single histidine residue and a seven conserved cysteine residues that are arranged in a C4- C2HC consensus sequence [26]. In addition to this, another feature is CTD (C- terminal domain) known as tetratricopeptide repeats (TPR) domain, a motif important in the binding of co-chaperones such as HSP90 is found in SMYD1-4 and is absent in SMYD5 which makes it different from other SMYD proteins [26]. Whereas, SMYD4 contains additional TPR domain at its N-terminus. SMYD protein family is known to methylate a variety of histone and non-histone proteins that contribute to a myriad of biological processes. Smyd5, a 47kD protein is identified as an H4 Lys20 methyltransferase and contains a polyglutamate insertion in its C-terminal domain. Our laboratory demonstrated that SMYD5 is localized in the nucleus which translocate from the nucleus to cytoplasm upon hydrogen peroxide treatment. Biological roles of SMYD5 have started to emerge in the area of immune reaction, embryonic stem cell regulation, hematopoiesis regulation, DNA double stranded break repair as well as cancer inhibition but at the same time, the physiological and biochemical function of SMYD5 remains largely unknown [25]. Therefore, it would be interesting to determine its role and function and gain further mechanistic insights.

1.1.7 KMTs in Budding Yeast:

Proteins undergo one of the major post-translational modification known as methylation. So far, KMTs methylating histone proteins contain the catalytic SET domain required for their enzymatic activity [28] whereas, only one histone KMT, Dot1, possesses a seven-β stranded domain in Eukaryotes [29]. Lysine 79 methylation of H3 occurs in eukaryotes ranging from budding yeast to human. K79 methylation in budding yeast is mediated by silencing protein DOT1 and H3 -K79 methylation is conserved from Yeast to human [28]. Budding yeast contains about 12 members of KMTs containing SET domain family, while the human proteome contains more than 50 KMTs [29]. Six methyltransferases involved in various biological processes were

examined in budding yeast namely, Rkm4, Rkm1, Efm7 and Efm4 which are known to methylate specific lysine or N-terminal residues of different subunits of translational machinery [30, 31]. Hmt1 is known to regulate post transcriptional processes [32]. SET5 which is the identified as first histone methyltransferase monomethylating the H4 at three lysine residues in the context of stress response [29]. It methylate's H4 tail at lysine 5, 8 and 12 in budding yeast [29]. SET5 and SET6, contains the split SET domain and two consecutive Zinc finger motifs. These structural features as well as the sequence comparison designates that SET5 is orthologous to human SMYD3 protein. SMYD3 is also known to methylate H4K5 which specifies functional conservation of the activity of SET5. The study of H4 methylation and SET5 in budding yeast may thus reveal conserved mechanisms that add to SMYD3-dependent oncogenesis and the evolutionary mechanism of chromatin function in human can also be known [29]. Until the present, HMT1 is the only methyl transferase in yeast whose biological structure has been solved [33]. So, our goal is to solve and study the structure and function of SET5 in budding yeast in order to understand the mechanism of methylation.

1.1.8 Ubiquitin Conjugating Enzyme:

One of the small regulatory proteins present in all eukaryotic cells is ubiquitin. It is a highly conserved protein and can exist either free or be covalently linked to other cellular or nuclear proteins [34, 35]. Ubiquitin conjugates with the intracellular proteins and marks their selective degradation. This process is known as ubiquitination [35]. Proteins are affected in many ways by ubiquitination. It can mark them for proteasomal degradation, alter their activity, change their cellular location etc. The three main steps in ubiquitination process are; Initially, the ubiquitin activating enzyme (E1) activates the ubiquitin molecule by binding to it. This activated ubiquitin molecule is then transferred to the ubiquitin conjugating enzymes (E2) which covalently attaches

the ubiquitin moiety to the target protein with or without the help of an E3 enzyme known as the ligase enzyme [36]. The protein is mark for proteasomal degradation when either single or more ubiquitin molecule is attached to it [37]. Therefore, the fate of the substrate i.e. signaling or proteasomal degradation is finally determined by ubiquitin conjugating enzyme (E2) [38]. A protein which catalyzes the covalent attachment of ubiquitin molecule to the specific proteins is encoded by a Rad6 gene [37, 38]. Rad6 protein which is highly conserved eukaryotic protein [39] plays an important role in various cellular processes including post-replication repair, damage-induced mutagenesis, sporulation, and DNA recombination [40]. It also plays a key role in proteasomal degradation pathway and cell signaling [38, 41, 42].The post-translational modification of targeted proteins via ubiquitin regulates the RAD6B pathway, which obligates them to rapid proteolysis [38, 40, 42]. This demonstrated that Rad6 protein is an essential ubiquitin-conjugating enzyme [37, 43] that plays an important role in maintenance of genetic integrity of mammalian cells. Imbalance in the levels of RAD6B protein could lead to chromosomal instability, aneuploidy, transformation and hypersensitivity towards drugs [42]. RAD6B is weakly expressed in normal human breast tissues but its expression rises in breast cancer cells [44]. Therefore, maintaining the critical level of RAD6B is very important [42]. Targeted treatment strategies using small molecule inhibitors are extensively studied [45]. SMI#9 is a selective small molecule inhibitor of RAD6B. My major goal is to solve the crystal structure of Rad6B in complex with SMI9 or its derivatives. Solving the structure of Rad6B/SMI9 complex will provide an excellent basis for rational design of more potent, selective Rad6B inhibitors.

CHAPTER 2: EXPRESSION, PURIFICATION, CRYSTALLIZATION AND ACTIVITY STUDIES OF HISTONE LYSINE METHYLTRANSFERASE: SMYD5

2.1 Introduction

SMYD protein family containing conserved SET and MYND domains plays a crucial role in myofibril assembly and is known to involved in muscle and heart development. This class of protein lysine methyl transferases has also been known to possess roles in tumorigenesis. So far, SMYD family consists of five paralogues. The catalytically active SET domain is responsible for cofactor binding, and methylation of lysine residues on histone and non-histone targets. The MYND domain is a zinc finger motif that plays a key role in protein-protein interaction.

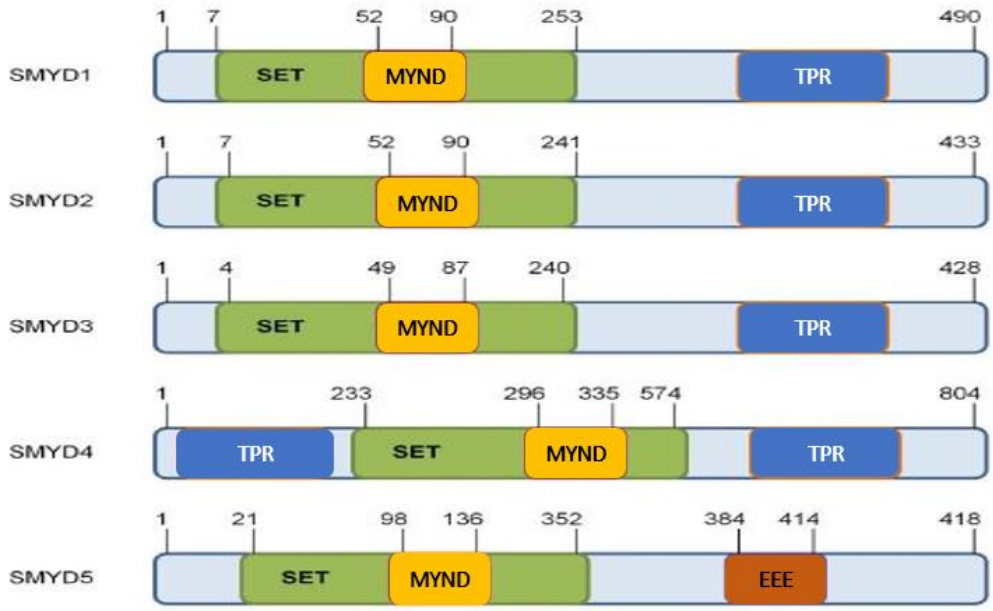


Figure 2.1: Overall structure of SET and MYND domain- containing protein family (SMYD) [46]

The diagram represents the structural domain organization of five members of SMYD protein family. The MYND, SET and TPR domains are highlighted yellow, green and blue respectively. Poly-E region of SMYD5 is highlighted brown. The numbers represent the

numbering of amino acids of each domain (numbers on top of each region) and overall size (numbers at the end) of the human proteins.

Research indicated that the structural and physiological roles played by the other SMYD members are distinct from that played by SMYD5. SMYD5 lacks TPR domain, important for the modulation of protein-protein interactions, and is present in other members of SMYD family. Instead SMYD5 has a poly-glutamate stretch. Unlike other members of SMYD family, SMYD5 does not have a function in the development of cardiac and skeletal muscle. SMYD5 was found to play a role in hematopoiesis in zebrafish. SMYD5 is associated with the NCoR co-repressor complex and was found to repress the expression of toll-like receptor 4 (TLR4) responsive genes in macrophages. It was also found to regulate pro-inflammation genes through trimethylation of H4K20. Embryonic stem (ES) cells self-renewal is directed by networks of epigenetic regulators and transcription factors. Disruption of these factors abolishes self-renewal. Recent studies indicated that, SMYD5 is an important regulator of ES cell function. SMYD5 is highly expressed in ES cells and its depletion resulted in compromised self-renewal. Deregulation of SMYD proteins has been informed to contribute to cancer progression through several different mechanisms. Research has indicated a possible role of SMYD5 in breast cancer through epigenetic alterations, most likely playing a role as tumor suppressor.

SMYD5, which is identified as H4K20 methyltransferase, may also play a larger role by methylating substrates other than histones. As stated earlier, the physiological function of SMYD5 is not well understood. Therefore, to understand its function better, one way is to identify its new substrates which will help us understand how SMYD5 recognizes its substrates.

2.1.1 Lysine Oriented Peptide Libraries (KOPLs) for Determination of KMT Substrate Selectivity:

Based on the substrate selectivity profile, our laboratory designed an artificial substrate peptide for SMYD5, known as KOPL peptide**.** The name of the peptide is based on the technique used to design it which is, "Lysine oriented peptide libraries". The peptide is 7 aa long, whose sequence is KAVKSYK. The substrate peptide screening for SMYD5 was done by our collaborators.

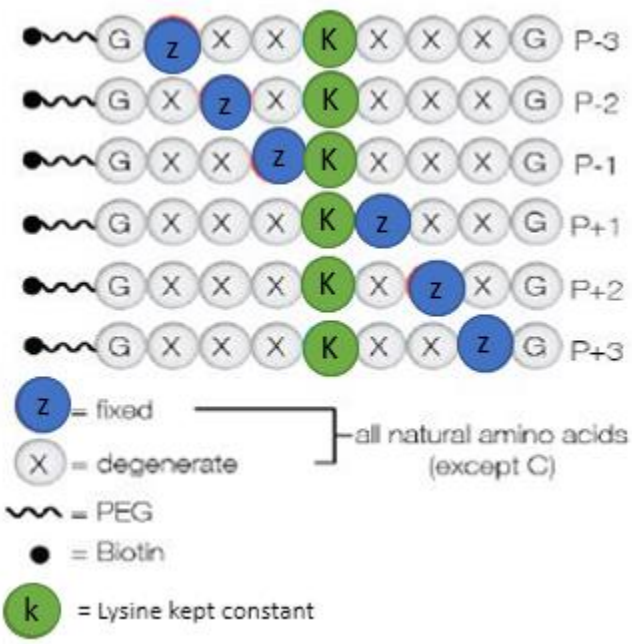


Figure 2.2: Screening result for KOPL peptide

In figure 2.2, 6 different positions (P-3 – P+3) were tested and in each position, the Z position is fixed which is one of the 20 amino acids. Xs are the degenerate amino acids which may vary from among the 20 amino acids. K (lysine) is always kept same. Accordingly, 20 amino acids will be tested for each position. The residue most preferred for each position was determined.

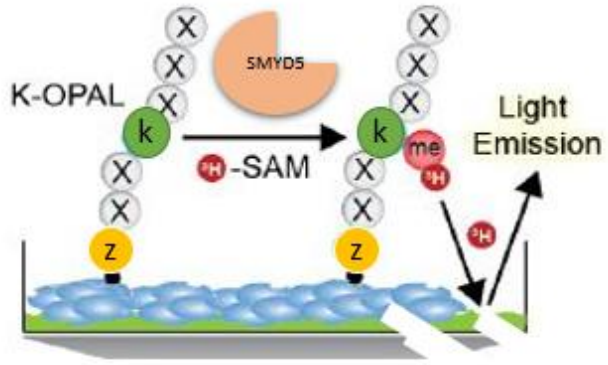


Figure 2.3: Schematic determination for determination of activity for each peptide with different sequences

Activity is performed in a 96 well plate, in which, the coated streptavidin binds to the biotinylated peptides, SAM is radiolabeled, KMT is SMYD5, me is the methyl group containing tritium group transferred from SAM (denoted in red); Light emission is measured to determine the methyltransferase activity.

In figure 2.2, radiolabeled SAM was used to detect the methyltransferase activity. Here, the activity was tested on each peptide at each position with different sequence. Once, the methyl group is transferred from SAM to the lysine residue by the methyltransferase, the amount of light emission is measured to determine the methyltransferase activity. The SMYD5 substrate selectivity profile can then be determined by the composite methylation score. Based on this, we designed the best substrate for SMYD5.

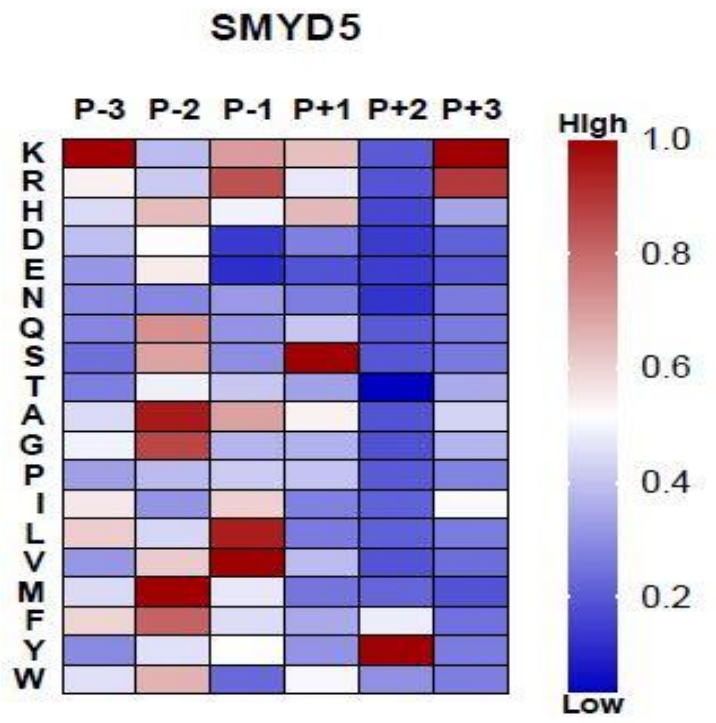


Figure 2.4: Composite score of each peptide with different sequences

Figure 2.4 is a heat map that represents the composite score for each position (P-3 – P+3). Some amino acids are better, and others are worse. In the left of the figure are all types of amino acids except for cysteine. To the top of the figure, are the 6 positions, and to the right of the figure, is the color scale which tells that, red means highly preferred residues and blue represents low preferred residues. From the map, we can say that, for the P-3 position K is the most preferred residue, For P-2 position M is the most preferred one. For P-1 V is the most preferred residue, K0 is target lysine, which is kept constant. S, Y and K are the preferred residues for P+1, P+2, and P+3 positions respectively. This is how we designed the KOPL peptide for SMYD5 whose sequence is **KAVKSYK**. Based on the substrate selectivity profile, the KOPL peptide is assumed to be the best peptide for SMYD5 and may facilitate its crystallization.

2.2 Materials and Methods

2.2.1 Protein Expression and Purification

The open reading frame of full length human SMYD5 was cloned into pCDF-SUMO vector with N-terminal His6-SUMO tag and transformed into the host BL21 (DE3) cells for recombinant protein expression.

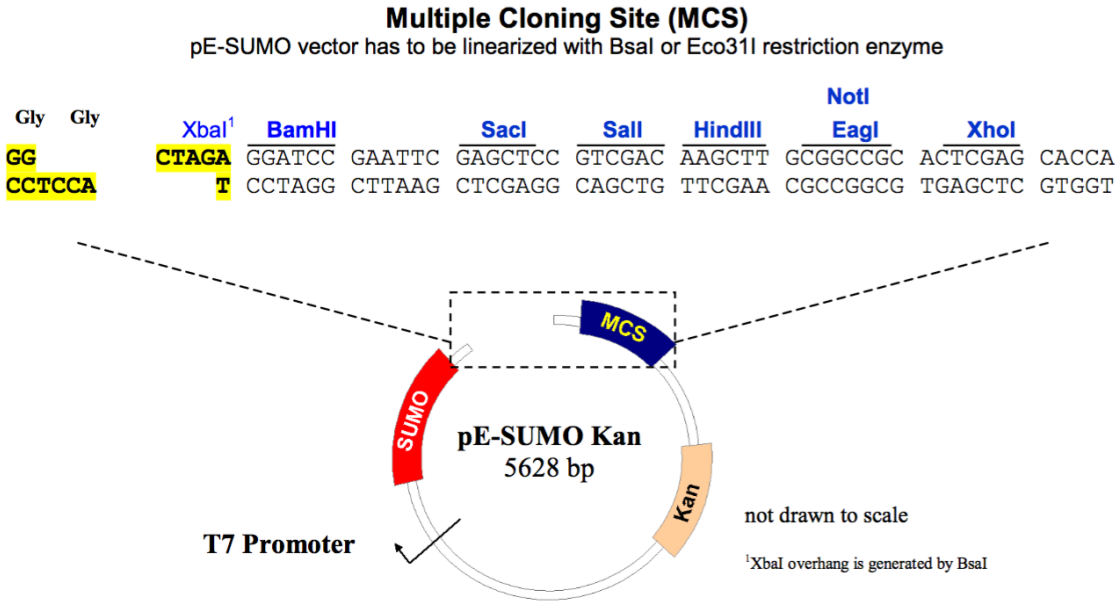


Figure 2.5: pCDF-SUMO Vector Map

2.2.2 Small Scale Expression Test

Small scale expression test is used to check whether the protein is soluble and expressed well. The conditions can be optimized accordingly. The transformants were grown in 5mL LB media to an optical density of 0.4-0.6 at 37ºC at 600nm. The culture was then induced with 0.1 mM IPTG and grown overnight at 15ºC. Cells were harvested next day and lysed by sonication. Induced and uninduced samples were collected to test expression, and cell lysate and supernatant

samples were collected to determine protein solubility. The results were analyzed on 15% SDS-PAGE.

2.2.3 Large Scale Expression

Large scale expression is used to express and produce the protein of interest for protein purification. The volume of cells to grow will depend upon (1) how well the protein is expressed, which can be analyzed from the small-scale expression test described above and (2) how much protein we want to have. The gene encoding for wild type SMYD5 protein was cloned into pCDF-SUMO vector and transformed into BL21(DE3) cells. Transformants were grown to an OD_{600} of 0.4-0.6 at 37º C in LB medium and protein expression was induced with 0.1mM IPTG (isopropylthio-β-D-galactoside). The cell growth was continued at 15ºC with shaking of 225 rpm overnight. Cells were harvested by centrifugation and suspended in binding buffer (20 mM sodium phosphate, 500 mM NaCl, 5% glycerol, 20 mM imidazole, 5 mM BME). Cells were lysed by French Press and the supernatant containing the soluble proteins was subjected to protein purification by three step chromatography using AKTA purification system including two times of Ni2+ affinity followed by size exclusion. The recombinant wild type SMYD5 contains a His6-SUMO-tag which was bound to HisTrap coloumn (GE Healthcare). The bound protein containing the His6 tag was eluted out by elution buffer (20 mM sodium phosphate, 500 mM NaCl, 5% glycerol, 500 mM imidazole, 5 mM BME). Fractions containing His6-SUMO-SMYD5 were pooled together. The SUMO tag bound to the protein was cleaved off by adding Ulp1 (SUMO protease 1). The native SMYD5 protein was collected by passing the protein through the second Ni2+ affinity chromatography column (HisTrap). Finally, SMYD5 was further purified through a Superdex 200 column (GE Healthcare) into gel filtration buffer (20 mM Tris pH 7.5, 150 mM

NaCl, 5% glycerol, 5 mM BME). Fractions containing pure wild type SMYD5 protein were pooled and concentrated and stored at -80ºC for further use.

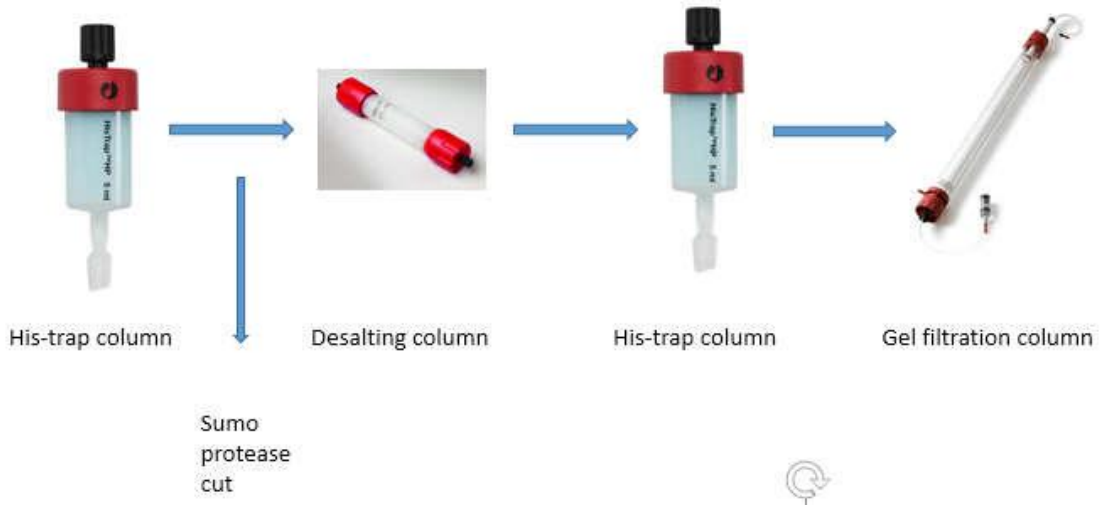


Figure 2.6: Overview of the purification strategy of SUMO-SMYD5 using AKTA system [47]

2.2.4 Methyltransferase Activity Assay

Methyltransferase assay is used to monitor the activity of a methyltransferase enzyme on its substrates. Here, we used a continuous kinetic assay to detect the methyltransferase activity of SMYD5. It is a robust assay for the continuous monitoring of methyltransferase activity and can be used to determine the steady-state kinetics (e.g., kcat and KM) for a variety of methyltransferases.

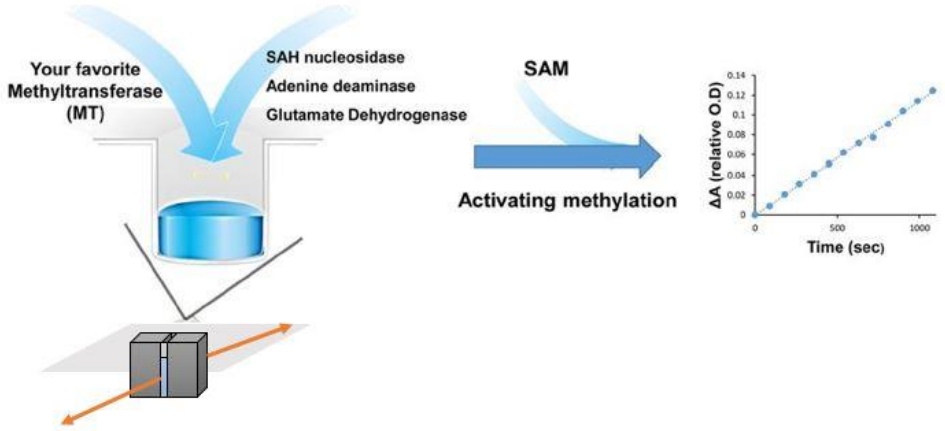


Figure 2.7: Continuous coupled assay for determination of methyltransferase activity [48]

Assay utilizes three enzymes: SAH nucleosidase (SAHN), Adenine deaminase (ADE) and Glutamate Dehydrogenase to couple methyl transferase activity to NADPH oxidation. SAM is used as methyl group donor. Activity was carried out in standard cuvette and monitored by a decrease in absorbance at 340nm.

In this assay, MTs activity is coupled to three different enzymes namely; adenine deaminase (ADE), SAH nucleosidase (SAHN) and glutamate dehydrogenase to couple methyltransferase activity to NADPH oxidation. As a result, the rate of NADPH oxidation reflects the rate of methylation and the methyltransferase activity can directly be monitored by a decrease in absorbance at 340 nm. The activities of SAHN and ADE are coupled with the activity of glutamate dehydrogenase that utilizes ammonia and alpha-ketoglutarate to generate glutamate followed by NADPH oxidation. The continuous change in absorbance at 340 nm due to NADPH oxidation is examined, which can be linearly correlated with the reduced concentration of SAM. Before the readings were documented, all the assay reactions were incubated at room temperature for 20 mins with the final volume of 60µl.

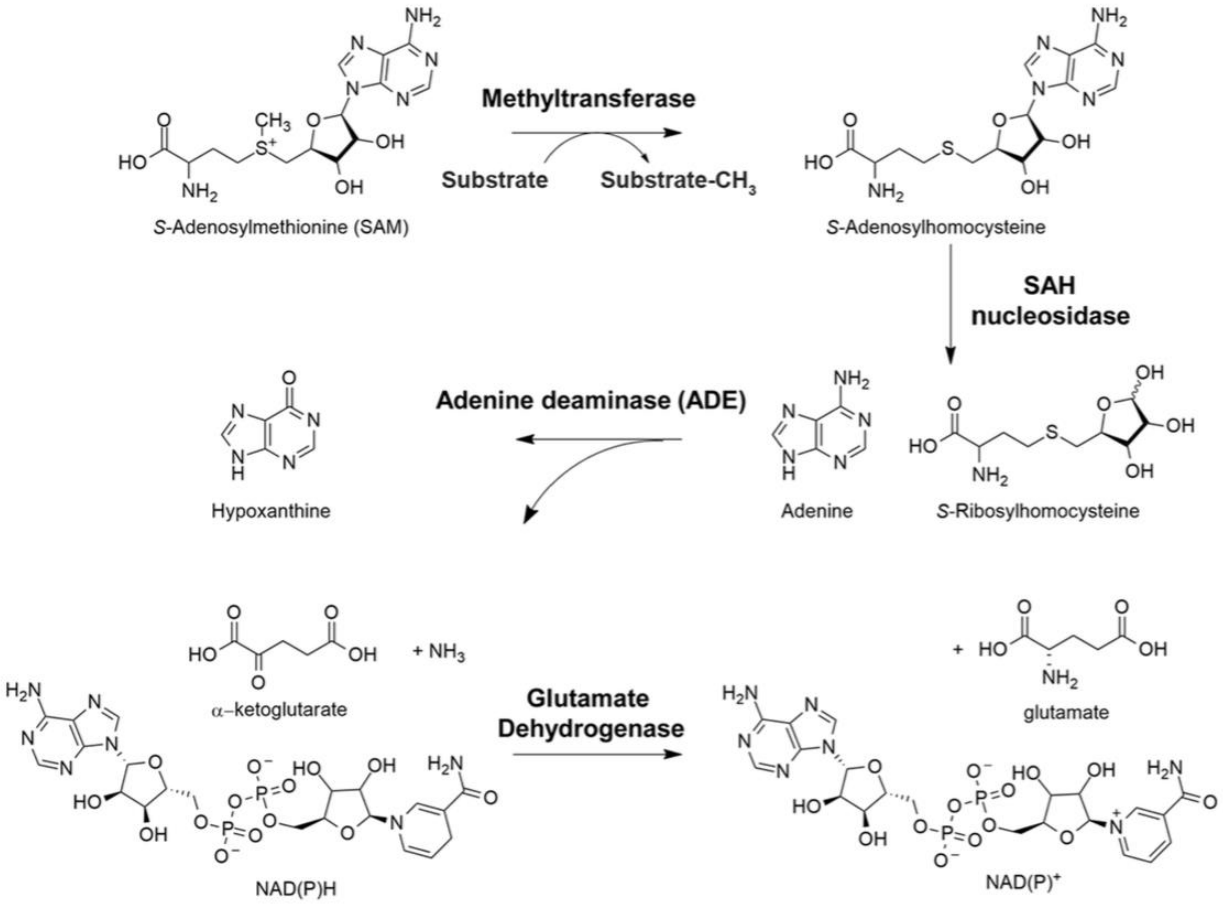


Figure 2.8: The chemical structure of all the products in methyl transferase coupled assay [48]

SAM is converted to SAH by methyltransferase enzyme. With the help of SAH nucleosidase, SAH is converted to Adenine + S-Ribosylhomocysteine. Adenine is then converted to hypoxanthine + α-ketoglutarate + NADPH by Adenine deaminase. α-ketoglutarate is then converted to glutamate and NADPH is oxidized to $NAD(P)^+$ by Glutamate Dehydrogenase.

2.2.5 DOT Blot Assay

A dot blot is a method used to detect peptides and proteins. It can also be adapted to determine the methyl transferase activity by detecting methylated products. It signifies a simplification of the western blot method, with that the proteins or peptides to be detected are not first separated by electrophoresis. Instead, the samples are applied directly on a membrane in a single spot, and the blotting procedure is performed.

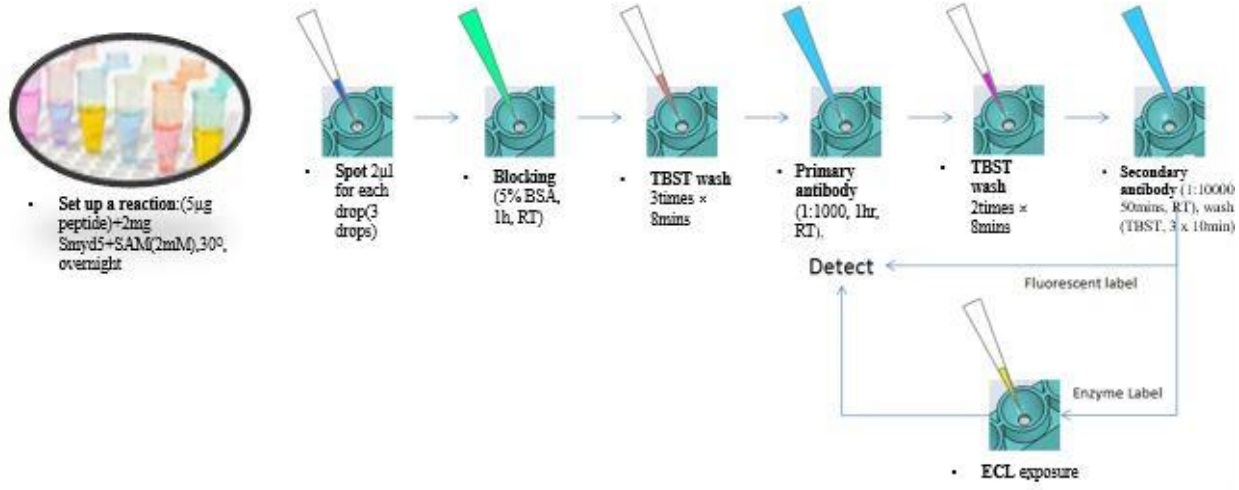


Figure 2.9: Overview of DOT Blot assay technique

Reaction was set up in 1.6 mL microcentrifuge tube at 30ºC overnight. The reaction contains 5µg KOPL+2mg/ml SMYD5+2mMSAM. Next day, 3 drops, 2µl of each, were plotted on nitrocellulose membrane. Blocking was carried out with 5% BSA for 1 hour at room temperature. Next, washing of membrane was carried out with TBST buffer for 3 times, 8 mins of each. Then, membrane was incubated with primary antibody (1:1000) for 1 hour at room temperature. Then, membrane was washed two times for 8 mins to remove any traces of primary antibody. Then the membrane was incubated with secondary antibody (1: 10000) for 50 mins followed by TBST wash of 3 times. Then, the secondary antibody labelled with HRP conjugate can be detected by ECL exposure.

2.2.6 Protein Crystallization of SMYD5

MBP-SMYD5(purified by one of the laboratory members) was used to set up the crystal screens. The buffer exchange was performed prior to crystallization, and SMYD5 was transferred into a bicine buffer (50mM bicine pH 9.0, 50mM NaCl, 0.5mM DTT). Protein concentration was diluted to 10 mg/mL. Various crystal screens such as, Crystal Screen 1 and Index Screen (Hampton Research) were used to perform crystallization. Crystal screens were also set up by incubating 10 mg/mL MBP-SMYD5 protein and 1 mM concentration of SAM for 1 hour on ice. Both hanging drop and sitting drop vapor diffusion techniques were used to set up crystal screens, by the addition of 1 ul of the well solution to 1 ul of protein sample. These crystallization trials were performed at 21°C.

2.3 Results

2.3.1 SMYD5 Crystallization Attempts

Full length SMYD5 gene was successfully cloned into pCDF-SUMO vector in the BL21 cells and was used for the future experiments. SMYD5 protein was expressed and purified by AKTA purification system by three step chromatography. About 1.25 mg of SMYD5 protein was purified from 4L of E.Coli cells.

Crystallization trials using MBP-SMYD5 were conducted with different sparse matrix crystal screens, were conducted at 21ºC. The sitting drop and hanging drop vapor diffusion techniques were used. Conditions appeared mostly heavily (Figure 2.12) or lightly precipitated (figure 2.10). Only 1 or 2 conditions showed phase separation (figure 2.11). However, no promising conditions were obtained.

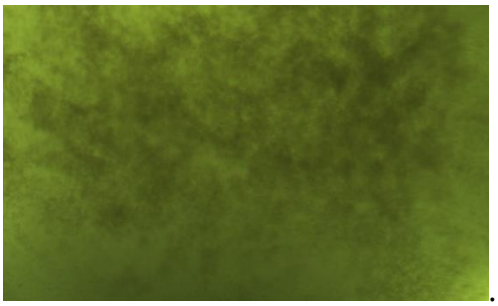


Figure 2.10: Light precipitated condition

0.2M sodium citrate tribasic dihydrate,0.1M sodium cacodylate trihydrate pH 6.5, 30% v/v 2-propanol

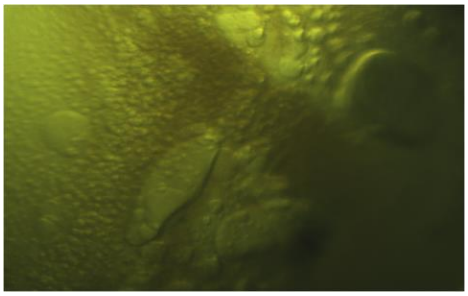


Figure 2.11: Phase separation condition

0.2M Ammonium sulfate, 0.1M sodium cacodylate trihydrate PH 6.5, w/v PEG 8000

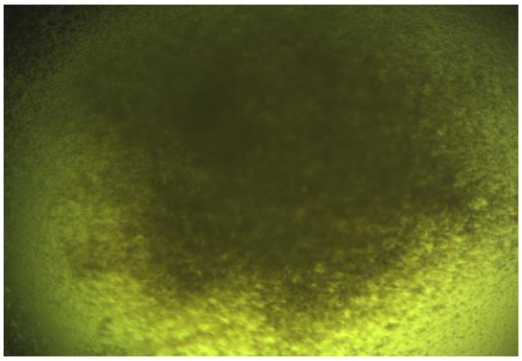


Figure 2.12: Heavy precipitated condition

0.2M Calcium chloride dehydrate, 0.1M HEPES sodium pH 7.5, 28% v/v PEG 400

2.3.2 SMYD5 Methylate's KOPL Peptide in Kinetic Assay

Kinetic assay was performed with a full length human SMYD5 in bicine buffer, pH 9.0. When all the assay components were present, a gradual change in absorbance at 340 nm was observed. This assay was used to measure the methyl transferase activity of SMYD5 on KOPL peptide with SAM as a methyl group donor. Two control reactions were run: (1) 500 µM NADPH, 4.5 µM SAHN, 3 µM ADE and Glutamate 2.62U (Figure 2.13) and (2) 300 µM KOPL peptide, 300 µM SAM, 4.5 µM SAHN, 3 µM ADE, Glutamate 2.62U (Figure 2.13). These control reactions ensured that the change in absorbance at 340 nm reflected the true measurement of methylation activity of SMYD5 on KOPL peptide. No activities were detected for the control reactions.

When the reaction was run, which contained 5 µM of full length SMYD5, 300 µM KOPL peptide, 300 µM SAM, 4.5 µM SAHN, 3 µM ADE and Glutamate 2.62U (Figure 2.13), a significant change in the absorbance at 340 nm was detected indicating the methyltransferase activity of full length SMYD5 protein on KOPL peptide by using SAM as a methyl group donor. As a result, the methyltransferase activity was directly monitored by the decrease in the absorbance at 340 nm.

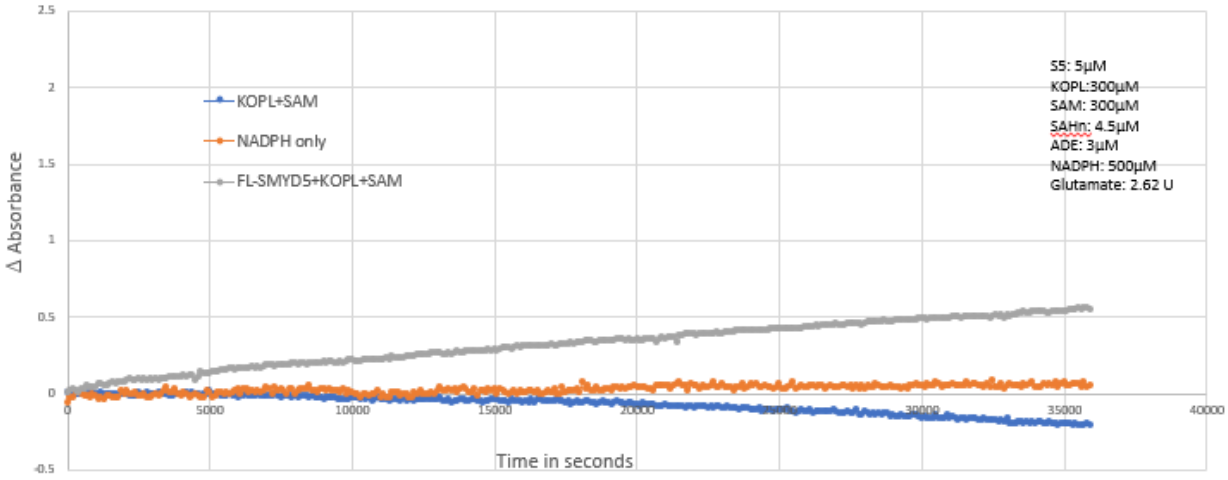


Figure 2.13: Methyltransferase activity monitored at 340nm

Blue line (KOPL+SAMNADPH+ coupled enzymes) and orange line (NADPH+ coupled enzymes) indicates the control reactions do not detect any activity. Grey line (SMYD5+KOPL+coupled enzymes) indicates the methyltransferase activity of SMYD5 on KOPL peptide by the change in the absorbance at 340 nm.

2.3.3 SMYD5 methylate's KOPL peptide in Dot Blot

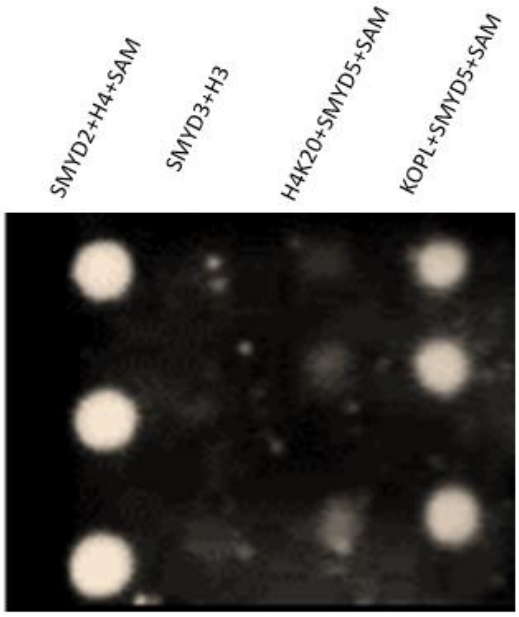


Figure 2.14: Dot Blot shows that SMYD5 methylate's KOPL peptide

H4 is a known substrate for SMYD2 and was used as a positive control in the Dot Blot assay. SMYD2+H4+SAM reaction gave positive signal. No signal was detected for SMYD3+H3 because, no SAM was added. It was used as a negative control in the reaction. For SMYD5, a little activity was detected with the H4K20 peptide, which is the known substrate of SMYD5; whereas, KOPL+SMYD5+SAM reaction gave a strong positive signal which suggested that SMYD5 could potentially methylate KOPL peptide.

2.4 Discussion

Our results from kinetic coupled assay and Dot Blot assay detected the methyltransferase activity of full length SMYD5 on KOPL peptide. These results were consistent with one another and suggested that, full length SMYD5 can methylate KOPL peptide, which is a synthetic artificial SMYD5 substrate designed by our laboratory. SMYD5 showed very little activity on H4K20 peptide, which is a known substrate for SMYD5.

In this study, multiple approaches were used to obtain the crystals. MBP-SMYD5 protein was used to perform crystallization trials. Crystallization conditions with different sparse matrix screens were performed. Yet, the trials yielded no crystals. One of the possible reasons may be due to the fusion tag. The fusion tags are known to increase the solubility of target protein and are being used for protein crystallization techniques to facilitate crystal lattice formation. However, despite increasing the chances of crystallization, the tagged proteins may have conformational heterogeneity due to the linker region and might not form crystals.

Crystallization conditions are specific to each protein. In future, crystallization experiments will be performed using full length SMYD5 in complex with KOPL peptide. Our results indicated that, full length SMYD5 can methylate KOPL peptide which may enhance the chances of crystallization. Second approach will be to perform re-cloning by designing new constructs that will exclude the flexible C-terminal region which might enhance the protein rigidity and facilitate crystallization of SMYD5.

In summary, this study provided evidence that KOPL peptide, which is designed based on SMYD5 substrate selectivity profile, could be the best substrate for SMYD5 protein. Alternative approaches like radioactive assays that involve the use of tritium (^{3}H) labeled SAM to study the methyltransferase activity can be used to further confirm that KOPL might be the first non-histone

substrate of SMYD5. In addition, this peptide can be used in co-crystallization experiments with full length SMYD5 that may facilitate its crystallization.

CHAPTER 3: TO SOLVE THE CRYSTAL STRUCTURE OF THE HISTONE KMTS SET5 FROM BUDDING YEAST (*SACCHAROMYCES CEREVISIAE*)

3.1 Introduction

Methylation of lysine residues is performed by lysine methyltransferases. SET5 is the first histone H4 lysine methyltransferase found in budding yeast that can add methyl marks to the lysine residues on histone H4. It is known to monomethylate the important H4 lysine residues 5, 8 and 12 in yeast.

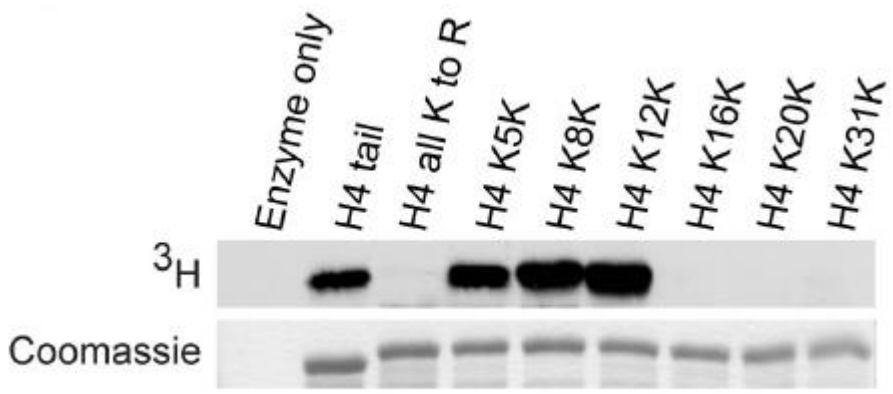


Figure 3.1: SET5 methylates lysines 5,8 and 12 [49]

It is also known that SET can methylate a native H4 in bulk purified histones and full length recombinant H4. It is also known to methylate lysine residues 4 and 7 on H2A to a minor extent. SET5 contains a split SET domain and has 2 consecutive Zn fingers. These structural elements and sequence comparisons specify that SET5 is orthologous to human SMYD3 methyltransferase, which likewise contains the split SET domain and zinc finger MYND domain. The study of SET5 and H4 methylation in yeast may therefore uncover conserved mechanisms that contribute to SMYD3 dependent oncogenesis in human cells.

Histone methyl marks indicate appropriate programming of the genome and have been associated in the advancement of several human diseases. Recent study suggested that SET5 might

interact with the chromatin to monomethylate histone H4 in cells. Yeast knockout assay for SET5 indicated that, loss of SET5 alone does not significantly affect global transcription. Genetic interaction studies suggested that SET5 functions with two global chromatin-modifying complexes namely; the COMPASS complex and NuA4 complex. COMPASS complex contains the H3K4 methyltransferase Set1 and the NuA4 complex contains the H4K5, H4K8 and H4K12 acetyltransferase Esa1. This data revealed that, H4 methylation and acetylation function in the same pathways [29].

The telomeres in budding yeast have been served as a model for studying the function of varied chromatin proteins in regulating and organizing chromatin structure and gene expression for many years. Expression of genes close to the telomeres in budding yeast are transcriptionally repressed by a combined set of chromatins modifying and remodeling elements. Apart from histone acetylation, histone H3 methylation also plays a role in silencing of adjacent genes to the telomeres. Previously, it was thought that silencing of genes adjacent to the telomeres was mainly mediated by the complex of Sir protein, but recent studies revealed the combined role of SET1 which is an H3K4 methyltransferase and SET5 which is known to methylate H4K5, H4K8 and H4K12 promotes a Sir protein-independent mechanism of telomere silencing [50]. Results stated a novel functional correlation amongst H3 and H4 methyltransferases, whose collective activities may possibly be associated in maintaining genomic integrity [51].

Additional biochemical and structural studies are required to reveal the mechanisms directing SET5 substrate selectivity and specificity and to understand the regulatory control of its enzymatic activity.

3.1.1 Protein Crystallization

The study of biochemical mechanisms and intermolecular interactions of a protein molecule can be performed by solving its macromolecular structure. This can be done either by nuclear magnetic resonance (NMR) or by using X-ray crystallography. NMR technique can only be used with low molecular weight proteins. X-ray crystallography is a technique used for determining the atomic and molecular structure of a protein which can be of higher molecular weight. X-ray diffraction patterns can be obtained from protein crystals. Getting a protein crystal is a difficult and an energetically unfavorable process. Achieving protein supersaturation is required for crystal formation, which requires manipulation of crystallization conditions such as pH, temperature, salt concentration, reducing agent, organic molecules etc.

In this method a protein crystal is exposed to an X-ray beam, resulting in a diffraction pattern collected on an X-ray detector. Then, by analyzing the angles and intensities of these diffracted beams, we can obtain an electron density map. From this electron density map the mean positions of the atoms of the protein can be modeled and determined, which finally leads to the 3D structure of the protein. (Figure 3.2)

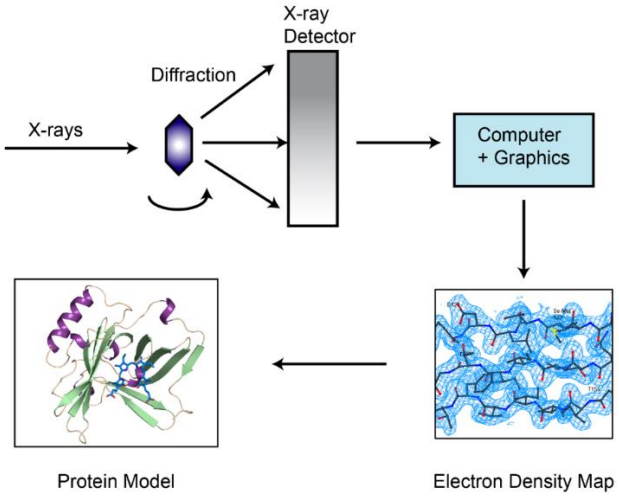


Figure 3.2: Overview of X-ray Crystallographic Method [52]

Out of all the methods known to crystallize a protein, the most commonly used technique is the vapor diffusion technique. Vapor diffusion takes place in a closed system, where the protein is mixed with the reservoir solution in 1:1 volume in a drop. A slow gradient of increasing protein concentration takes place as the solution evaporates from the drop and finally ends up into the reservoir to achieve the equilibrium. The two most common techniques used for vapor diffusion are sitting drop and hanging drop vapor diffusion techniques (figure 3.3) which shares the same principle.

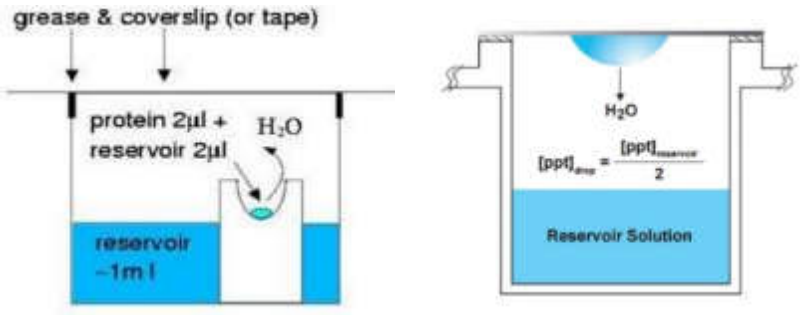


Figure 3.3: X-ray crystallographic methods [47]

Left represents sitting drop vapor diffusion technique. Right represents hanging drop vapor diffusion technique.

Once the promising conditions are obtained, they will be optimized further in order to get larger crystals and achieve better X-ray diffraction. Careful observation of the results of crystallization experiments is necessary to determine the parameters that are optimal for crystallization.

3.2 Materials and Methods

3.2.1 Crystallization

Sequence alignment of SET5 with the other members of the SMYD family revealed that SET5 is an orthologue of human SMYD3. However, obtaining crystal structure of SET5 will reveal how SET5 recognizes its substrates and how it is associated with human SMYD protein.

Purified His-SET5 was collected from our collaborators. Crystallization attempts were performed with the protein to identify conditions yielding diffraction quality crystals. The protein was diluted to 5mg/mL concentration. Sinefungin (SFG) was added in 1:10 volume of protein and cofactor. The protein and cofactor together were incubated at 4ºC for ~ 2 hrs. The protein mix was centrifuged at 11800 rpm for 10 mins to get rid of any aggregates that might prevent the formation of crystals. Sitting drop vapor diffusion technique was used to test different reservoir solutions to determine which conditions facilitate crystallization. Commercially available sparse matrix crystallization screens such as Crystal Screen I and Index Screen (Hampton Research) were used. Crystal screens were set up by addition of 1 µl of the well solution to 1 µl of protein mix and incubated at 22° C. The screens were monitored for optimized conditions.

3.3 Results

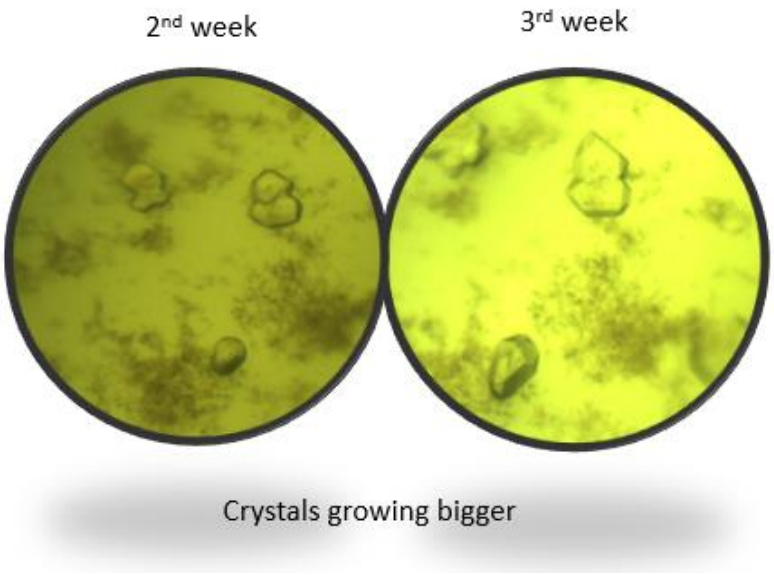


Figure 3.4: Set5 crystals

With the protein concentration of 5 mg/mL and crystal condition of 0.2 M ammonium acetate, 0.1 M Tris-HCL pH 8.5, 15% 2- propanol, a few crystals appeared within 2 weeks and continued to grow to its maximum size the following week. His-SET5 crystals were harvested and

flash frozen in liquid nitrogen. Crystals were sent to APS (Advance Photon Source) for X-ray diffraction data collection.

X-ray diffraction data from crystals will be collected at the Advanced Photon Source (APS) using LS-CAT X-ray synchrotron. Wayne State is a standing member of APS LS-CAT and has guaranteed access to synchrotron beam lines. APS offers the brightest X-ray beam that allows us to determine X-ray structures to the highest possible resolution as well as the ability to work with difficult samples. Once the data is collected, the structure can be solved by the molecular replacement method.

3.4 Discussion:

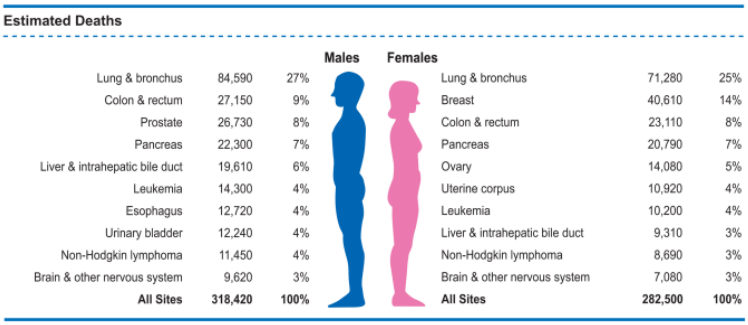
In this study we were able to find out the conditions that can obtain of SET5 crystals. A few crystals were obtained by the sitting drop vapor diffusion technique by mixing 5 mg/mL protein with the equal volume of reservoir solution containing 0.2 M Ammonium acetate, 0.1 M Tris-HCL pH 8.5, 15% 2- propanol.

In summary, this study provided the starting condition for optimization of SET5 crystals. This condition could potentially lead to obtaining well diffracted SET5 crystals. Attaining high resolution structures of SET5 would be of value since SET5 is a known orthologue of human SMYD3 protein. SMYD3 has been implicated in tumorigenesis. Hence, further studies on SET5 may uncover conserved mechanisms that contribute to SMYD3 oncogenesis in human cells.

CHAPTER 4: STRUCTURE BASED OPTIMIZATION OF THE INHIBITORS OF RAD6B, A POTENTIAL DRUG TARGET FOR BREAST CANCER TREATMENT

4.1 Introduction

Cancer is one of the most serious health threats and leading causes of death [53]. Breast cancer is an invasive cancer which generally affects women. It is the second main leading cause of deaths [54]. One in every 8 women will develop breast cancer in her lifetime. Many mutations occur as a result of problems caused by DNA lesions during replication [55]. This process termed as mutagenesis is a major pathogenic factor because the mutations are thought to inactivate cellular defenses against uncontrolled proliferation and cell migration which is a common factor in cancer. There is currently no cure for metastatic breast cancer, and this poses the need for strong cancer research which will enhance our understanding about cancer and help us to develop novel and more effective treatment techniques. Traditional therapies have been proved powerful in the treatment of cancer but one of its disadvantages is that it leads to adverse side effects due to its toxicity on both tumorous as well as normal cells [56, 57]. Targeted treatment strategies using small molecule inhibitors are being widely studied [58]. Small molecule inhibitors (SMI) are the compounds $\leq$ 500 Da in size. They can reach the intracellular target sites by diffusing across the plasma membrane due to their small size. These compounds can disrupt protein-protein interactions and can be used in cancer treatment [58].



Estimated Deaths

Males			Females		
Lung & bronchus	84,590	27%	Lung & bronchus	71,280	25%
Colon & rectum	27,150	9%	Breast	40,610	14%
Prostate	26,730	8%	Colon & rectum	23,110	8%
Pancreas	22,300	7%	Pancreas	20,790	7%
Liver & intrahepatic bile duct	19,610	6%	Ovary	14,080	5%
Leukemia	14,300	4%	Uterine corpus	10,920	4%
Esophagus	12,720	4%	Leukemia	10,200	4%
Urinary bladder	12,240	4%	Liver & intrahepatic bile duct	9,310	3%
Non-Hodgkin lymphoma	11,450	4%	Non-Hodgkin lymphoma	8,690	3%
Brain & other nervous system	9,620	3%	Brain & other nervous system	7,080	3%
All Sites	**318,420**	**100%**	**All Sites**	**282,500**	**100%**

Figure 4.1: Cancer death statistics [53]

Rad6 gene encodes a highly conserved eukaryotic protein, RAD6B, that plays a role in the covalent attachment of ubiquitin to the other targeted proteins. RAD6B protein is an important E2 conjugate that mediates ubiquitination reaction. This protein plays a key role in various cellular processes such as post-replication DNA repair, recombination and damage induced mutagenesis.

One of the substrates for RAD6B is histone H2A protein. Ubiquitination of histone H2A is one of the post translational modification in DNA damage response. Ubiquitination reaction involves three main steps: activation, conjugation, and ligation, performed by three different enzymes namely, ubiquitin-activating enzymes (E1s), ubiquitin-conjugating enzymes (E2s), and ubiquitin ligases (E3), respectively.

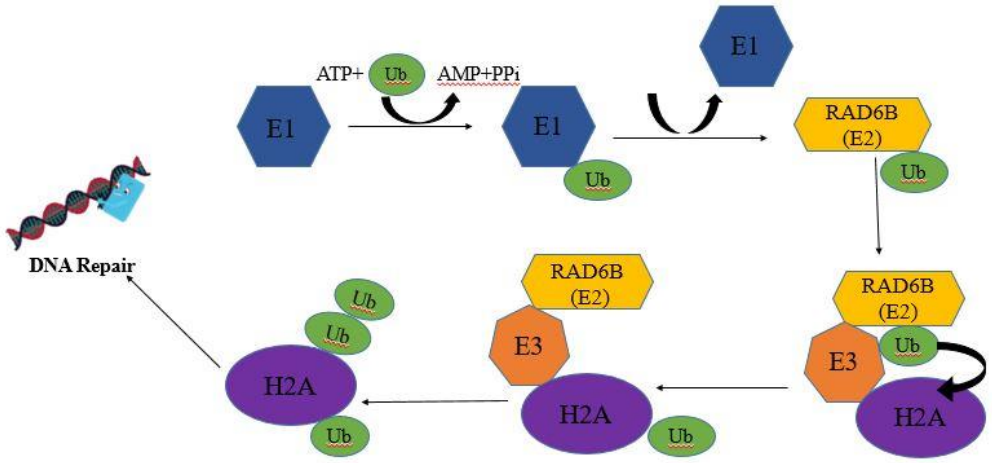


Figure 4.2: Overview of ubiquitination reaction mediated by RAD6B

Ubiquitin molecule is activated by ubiquitin-activating enzyme (E1) which is ATP dependent step. Next, ubiquitin-conjugating enzymes (E2) catalyzes the transfer of ubiquitin from E1 to the active site of the E2. Ubiquitin ligases (E3), catalyzes the final step of the ubiquitination cascade. This reaction may lead to mono, multi or poly ubiquitination at the substrate molecule which finally leads to DNA repair.

Normal human breast tissues weakly express Rad6B protein but its expression increases in breast cancer tissues [44]. Thus, deregulation in expression of RAD6B may be an important step in transformation to the malignant phenotype [42, 44]. β catenin or other components of Wnt pathway are activated in most breast cancers. An increase in the level of β catenin, states its crucial role in breast cancer[58]. Rad6B which is overexpressed in breast cancerous cells plays a remarkable role in β catenin ubiquitination and stabilization [59]. Therefore, maintaining the critical level of RAD6B is very important.

Aiming at Rad6B would be a promising way for breast cancer treatment. Proteasome inhibitors which are inaccessible to provide specificity will be achieved by small molecule inhibitors (SMI) which are selective for Rad6B [60]. The human Rad6B protein has been crystallized previously [61]. With the help of molecular docking, the novel SMIs #8 and #9 that interfere with Rad6B catalytic sites have also been identified [60]. SMI#9 more effectively induced apoptosis, cell cycle arrest in G2-M phase and inhibited proliferation, migration and colony formation and Rad6B induced histone ubiquitination. SMI#9 is also known to decline β catenin levels [60].

The crystal structure of Rad6B in complex with SMI#9 is not available. This has prevented a rational design of inhibitors with better selectivity and potency. Solving the crystal structure of Rad6B/SMI#9 complex will provide us with more accurate binding information of the key residues

of Rad6B which interacts with its small molecule inhibitor SMI#9. This will provide us with the information of the best binding poses of SMI#9 with Rad6B. Based upon the structural findings, the compound can be optimized in terms of electrostatic interactions, the fit in ligand binding pocket, and van der Waals contacts. This data is expected to create new Rad6B inhibitors with enhanced potency and selectivity. The generated compounds will be drug-like, synthesizable and can be used to inhibit the overexpression of Rad6B in breast cancer and hence can be applicable in breast cancer treatment.

4.2 Materials and Methods

4.2.1 Molecular Cloning and Protein Expression

RAD6B gene was initially present in PET101 vector which was collected from our collaborators. The gene of interest needs to be amplified from genomic or vector DNA by PCR (polymerase chain reaction) before it can be cloned into an expression vector. The first step is to design the necessary primers.

5’ primer was designed which was 35 bp in length, TTATATGGTCTCAAGGTATGTCGACCCCGGCCCGG (Tm=68ºC) and included the Bsa1(GGTCTC) restriction enzyme site. 3’ primer was 33bp in length, GCCGCGCTCGAGTTATGAATCATTCCAGCTTTG (Tm=65.1ºC) and included a Xho1 restriction site. The primers were ordered from Invitrogen company.

DNA fragment amplification encoding full length Rad6B was performed by PCR and then cloned into the pCDF-SUMO vector. RAD6B gene was amplified by PCR and then analyzed on a 0.8% agarose gel (Figure: 4.3). This was followed by the purification of PCR products using the PCR cleaning kit (Qiagen company) and then performing double digestion using Xho1 and Bsa1

restriction enzymes. Vector and insert digested by the restriction enzymes were purified and extracted using gel extraction kit (Qiagen company).

Purified insert and vector were mixed together with quick DNA ligase enzyme to circularized into a vector DNA and then followed by transforming the vector DNA into DH5α cells for plasmid expression. Plasmid purification was performed using GenEluteTM Plasmid Miniprep kit (Sigma), and then plasmid samples were sent for DNA sequencing to GENEWIZ.

4.2.2 Transformation for Protein Expression

1 µL of miniprep product containing the vector DNA was transformed into the 50 µL of codon optimized BL21 cells and heat shocked for 45 seconds at 42˚C. 250 µL of prewarmed SOC medium was added to the BL21 cells and incubated for 1 hour with 225 rpm of shaking at 37 ˚C and then plated onto streptomycin LB agar plates to get the recombinant colonies. Glycerol stocks were made with the overnight 5mL culture made from a single recombinant colony. 1 mL BL21 cell culture+300µL of 80% glycerol was mixed together in 2mL coring tubes, labeled and stored at -80˚C for future use.

4.2.3 Small Scale Expression Test

For recombinant protein expression, BL21 codon plus cells containing RAD6B gene were grown at 37ºC to an optical density of 0.4 – 0.6 and protein expression was induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were grown overnight at 15ºC with 225 rpm of shaking. Next day, the OD600 was measured, and cells were lysed using sonicator. Lysate and supernatant were collected to compare for the solubility test of a protein. The results were analyzed on 15% SDS PAGE. (Figure 4.4)

4.2.4 Large Scale Protein Production by Induction with IPTG

5mL LB media +1X streptomycin was inoculated with the BL21 cells containing RAD6B gene in pCDF-SUMO plasmid. Overnight culture was added to 500mL prewarmed LB medium containing streptomycin antibiotic for recombinant RAD6B protein production and purification. The sample of overnight culture was saved for SDS-PAGE analysis. The 500 mL inoculated culture was grown at 37° for 2 – 3 hours until the OD600 was between 0.4 – 0.6. The culture was slowly cooled down to 15ºC and induced with 0.1mM IPTG and grown overnight. The induced sample was saved for SDS-PAGE analysis. (Figure 4.5)

4.2.5 Protein Purification by AKTA System

Rad6B was purified by three-step chromatography including two times of Ni2+ affinity followed by size exclusion. The cells grown by IPTG induction were harvested by centrifugation and lysed by French Press. The soluble fraction was then subjected to Ni2+ affinity chromatography purification. The His-SUMO tag was removed by SUMO protease1 to produce native RAD6B. RAD6B proteins were separated from the cleaved tag by a second Ni2+ affinity chromatography and further purified by size-exclusion chromatography. Finally, the proteins were concentrated to 10–20 mg/mL in a buffer containing 20 mM Tris–HCl (pH 7.5), 150 mM NaCl, 1 mM β-mercaptoethanol, 5% glycerol, and 5 mM BME. Fractions containing pure RAD6B were pooled, concentrated to 35.4 mg/mL and stored in the -80°C for future use. (Figure 4.6-4.10)

4.2.6 Western Blotting

Purified RAD6B protein with three different concentrations (0.5µg, 1µg, 2 µg) were loaded onto 15% SDS-PAGE for electrophoresis and then transferred onto a nitrocellulose membrane. Next, membrane blocking was performed using nonfat dry milk for 1 hour at room temperature followed by 5 minutes of 1X TBST wash for 3 times. Immunoblotting was performed by

incubating the primary antibody (1:3000) against RAD6B, overnight at 4°C. The membrane was then washed 3 times for 5 mins using 0.1% TBST and the secondary antibody (1:5000; Goat anti-mouse IgG(H+L) HRP, Invitrogen) was incubated for 1 hour at room temperature. ECL (HyGLO HRP detection kit) was then used to detect the signal. The blots were developed using X-ray film where the light is emitted from the chemical reaction between secondary antibody, which is conjugated to HRP and the ECL solution used to detect the signal.

4.2.7 Activity Assay

Activity assay was carried out where, 1 µg of RAD6B protein bought from Boston Biochem laboratory was used as a control and three different concentrations, 1 µg, 2 µg and 4 µg of purified RAD6B in our laboratory were used to test the activity. All the proteins were incubated with histone H2A (1 µg), ubiquitin (0.5 µg), ATP/Mg^{+}2 and E1 enzyme in 10x UBC buffer for 1 hour at 30ºC. The reaction was stopped by boiling with SDS buffer containing BME and analyzed by western blot, same as described before.

4.2.8 Crystallization

Crystallization was performed using the hanging drop vapor diffusion method with the purified RAD6B. Various crystal screens were set up with and without SMI#9. The crystals were obtained by mixing the equal volume of reservoir solution containing 0.1M bisTris, pH 7.5,0.1M sodium tartrate, 1.3 M ammonium sulphate, 10% DMSO and RAD6B protein (10mg/ml) at 20ºC. Crystals appeared in one day and continued to grow for next 2 to 3 days. Crystals were stained with 0.1% of methylene blue to confirm if the crystals obtained were protein crystals and not salt crystals. The crystal appeared blue after staining indicating protein crystals.

4.3 Results

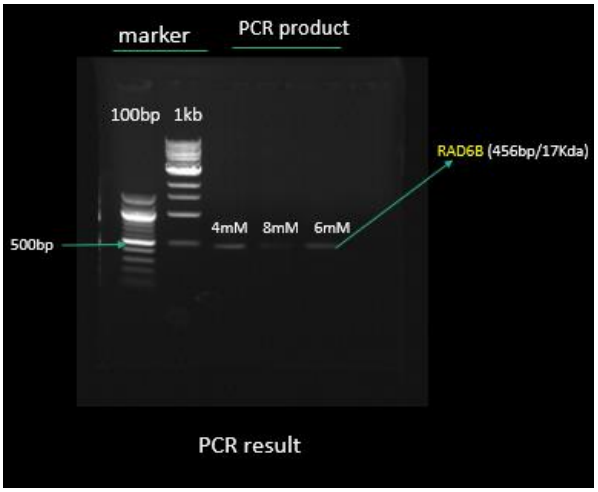


Figure 4.3: PCR result

RAD6B gene inserted in pCDF-SUMO vector was amplified by PCR and analyzed on 0.8% agarose gel. The 1st and the 2nd lane represent the 100bp and 1 kb DNA markers. Lanes 3rd, 4th and 5th represents the band corresponding to the molecular weight of RAD6B gene which was amplified by PCR with 3 different concentrations of MgCl2, 4 mM, 8 mM and 6 mM respectively.

4.3.1 RAD6B Construct Cloning and Purification

The RAD6B gene was transformed into BL21 cells for protein expression and purification. Small scale expression test was performed to check the solubility and expression level of recombinant RAD6B.

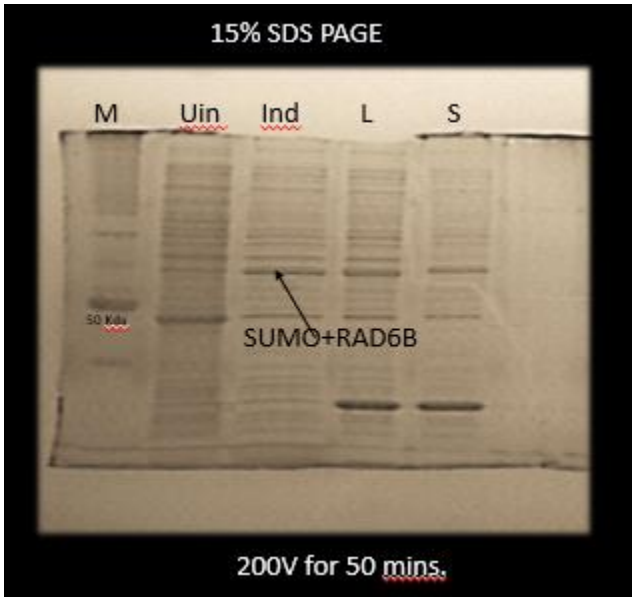


Figure 4.4: Small scale expression test

In the figure (4.4), RAD6B protein containing the SUMO tag is indicated with the corresponding band of expected molecular weight. The 1st lane represents protein marker. 2nd and 3rd lane represents uninduced and induced samples. 4th and 5th lane represents lysate and supernatant samples.

RAD6B protein purification was performed using AKTA purification system. In large scale purification process, the RAD6B expression and solubility tests were further confirmed by running 15% SDS PAGE. The cell lysate containing the SUMO tagged RAD6B protein was loaded onto the 1st HisTrap column.

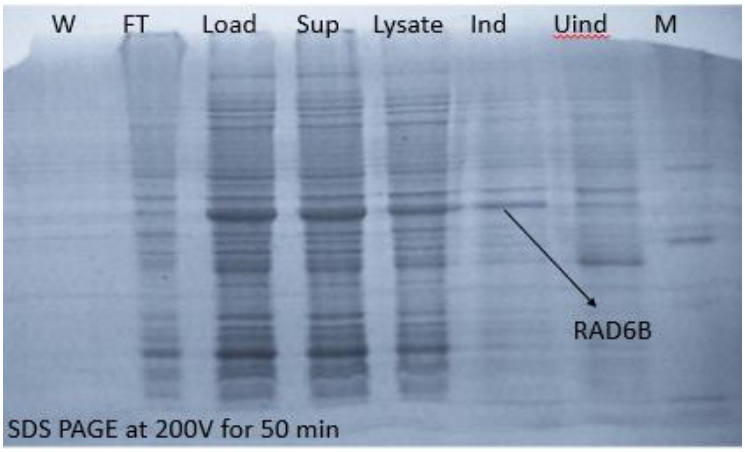


Figure 4.5: Large scale expression test

1st, 2nd and 3rd lanes represent the wash, flow through and sample loaded onto the column. 4th, 5th, 6th and 7th lanes represent the supernatant, lysate from French pressed cells, induced and uninduced samples from large scale expression. The last lane is the protein molecular weight marker. SUMO-RAD6B band corresponding to its molecular weight is indicated.

According to the chromatogram of 1st HisTrap purification (Figure 4.6), A6-B2 protein fractions were analyzed by 15% SDS-PAGE (Figure 4.7), indicating that A8 to B1 contained most of the RAD6B proteins. These fractions were then pooled together and incubated with SUMO protease overnight to cut the SUMO tag from RAD6B, followed by desalting (Figure 4.8).

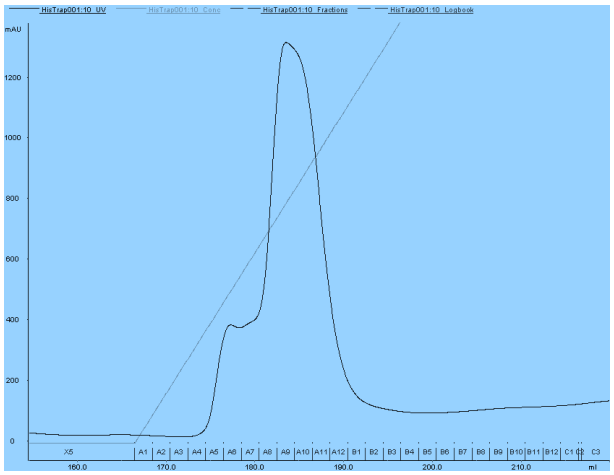


Figure 4.6: 1st HisTrap chromatogram

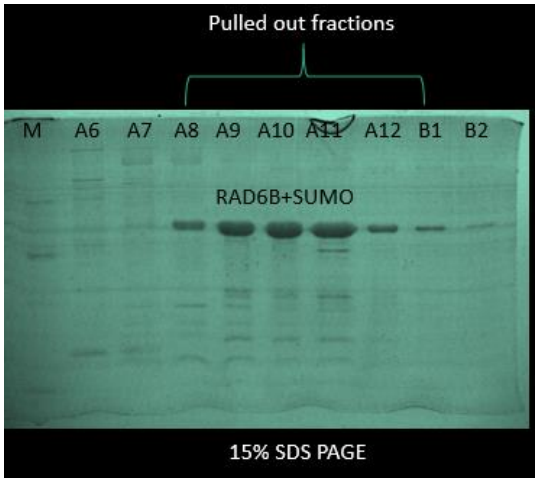


Figure 4.7: 1st HisTrap Fractions

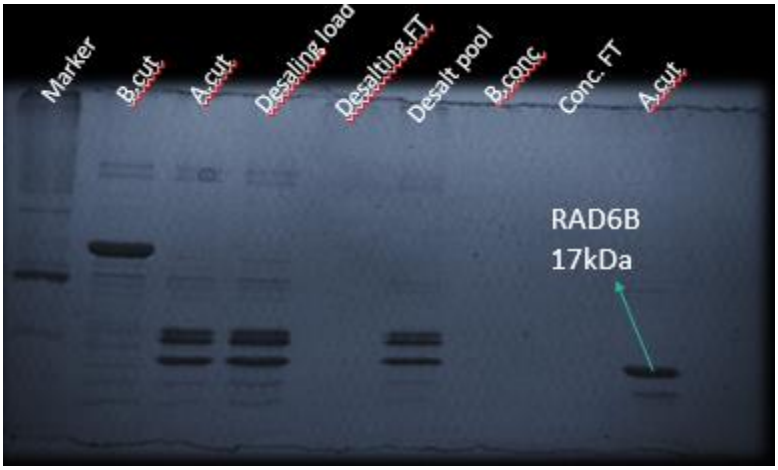


Figure 4.8: Loading protein to the desalting

Protein marker, before cut (B. Cut), after cut (A. Cut), desalting load, desalting flowthrough (FT), desalting pool, before concentration (B.conc), and concentrated flowthrough (Conc. FT), were analyzed by 15% SDS PAGE.

An aliquot of the samples from each step was saved and examined by running 15% SDS-PAGE (Figure 4.8). After running the samples on 2nd HisTrap column, SUMO tag was bound to the column and Rad6B was collected, concentrated and loaded onto Superdex 200 column. The fractions from the 2nd peak (figure 4.9) were collected and analyzed on 15% SDS-PAGE (Figure 4.10) in which C3- C7 fractions were pooled together, filtered and concentrated. Finally, 12 mg (35.4 mg/mL, 250µl) RAD6B protein was obtained from 2 liters of E.coli cell culture.

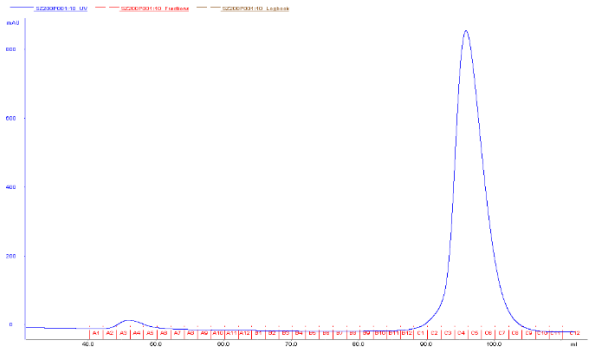


Figure 4.9: S200 chromatogram

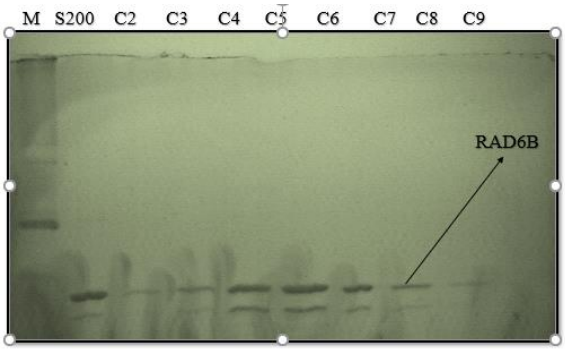


Figure 4.10: S200 fractions

4.3.2 Western Blotting

To confirm whether the purified protein was a RAD6B protein, western blot analysis was performed. Three different concentrations of the protein which was purified in our laboratory using AKTA purification system were analyzed. The results indicated that the protein was recognized by anti- RAD6B antibody and exhibited a band corresponding to the expected molecular weight (17 kDa) of RAD6B. This indicated that the purified protein was indeed a RAD6B protein.

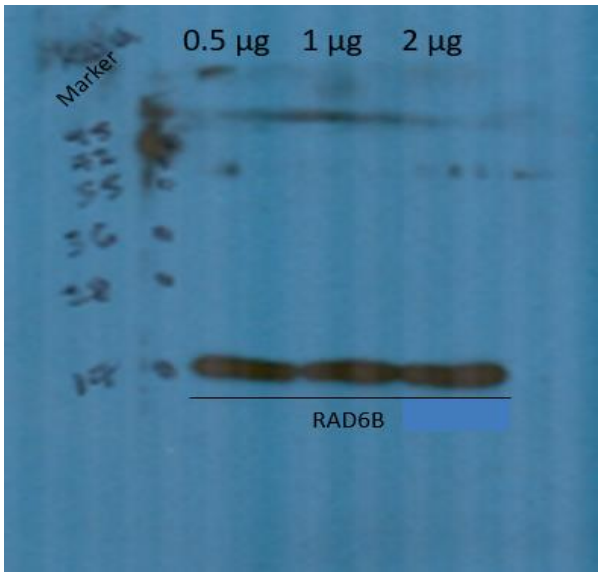


Figure 4.11: Western Blot analysis to detect the presence of RAD6B protein

The 1st lane is the molecular weight marker followed by three different concentrations (0.5µg, 1 µg, 2 µg) of purified Rad6B. The bands were seen at the molecular weight of 17kDa in all three concentrations.

4.3.3 Activity Assay

Activity assay was performed using anti-ubiquitin antibody to detect the ubiquitination reaction carried out by RAD6B protein on H2A. Here, we used three different concentrations of RAD6B protein purified in our laboratory (1 µg, 2 µg and 4 µg) to test the activity. 1 µg of the protein bought from Boston Biochem laboratory was used as a control. In Figure 4.12, bands corresponding to 10 kD and 17 kD marker were observed in all the four concentrations of protein indicating ubiquitin monomer and ubiquitin dimer respectively. In 2 µg and 1 µg (control) protein concentration only a single upper band was detected corresponding to 28 kD marker which suggested that it was a RAD6B protein containing a ubiquitin molecule. However, in the 4 µg protein concentration, two bands (upper and lower) around 28 kD marker were observed. The two bands were close to each other and suggested that the upper band was RAD6B with ubiquitin molecule attached to it while the lower band was H2A, to which the ubiquitin molecule was attached. This suggested that the Rad6B protein was in the process of transferring the ubiquitin molecule to its substrate H2A. This activity was observed clearly in 4 µg concentration of the protein and very little in 1 µg (control) concentration.

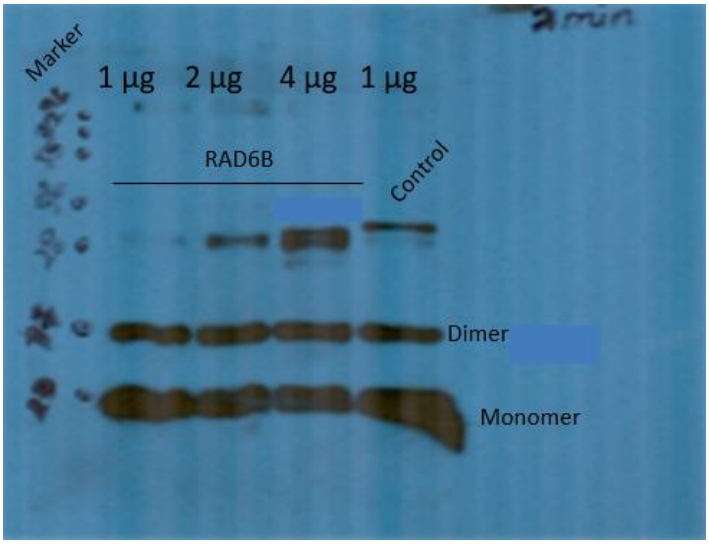


Figure 4.12:Ubiquitin conjugating activity of RAD6B

1 µg, 2 µg and 4 µg concentrations are the RAD6B proteins purified in our laboratory. The second 1 µg concentration is the protein bought from Boston Biochem laboratory.

4.3.4 Crystallization

Crystal screens were set up with purified RAD6B protein to obtain crystals. After introducing different concentrations of SMI#9 with different crystallization conditions. Crystals appeared the next day with 1.3M ammonium sulfate,0.1M bis-Tris pH 7.5,0.1M sodium tartrate, and 10 % DMSO. The crystals grew bigger the next day. The crystals appeared blue after addition of 0.1% of methylene blue which suggested the crystals were protein crystals. These crystals will be optimized to achieve adequate diffraction quality for data collection.

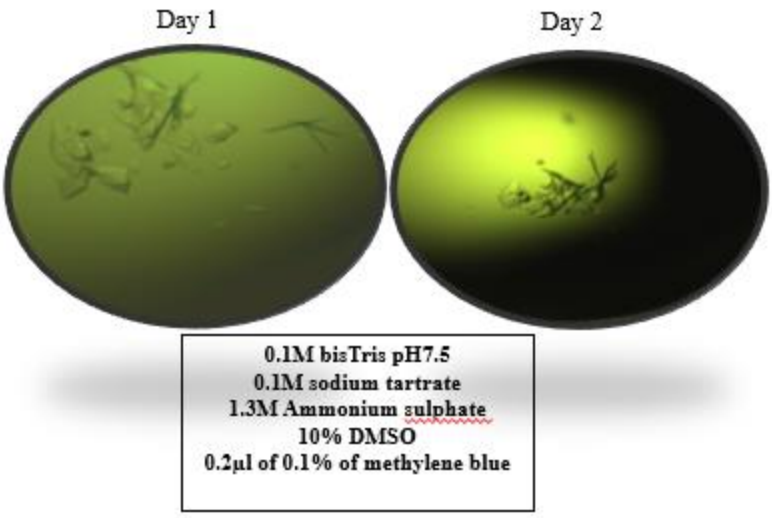


Figure 4.13: Crystals obtained for RAD6B protein.

Crystals obtained by hanging drop vapor diffusion technique using10mg/mL of protein. Crystals appeared overnight and continued to grow.

4.4 Discussion:

In this study, RAD6B gene was successfully cloned into pCDF-SUMO vector and a significant amount of proteins was purified using E. coli as a production host. We performed the western blot analysis of the purified protein using anti-RAD6B antibody. The results indicated that the protein we purified was indeed a RAD6B protein. Next, we performed an activity assay to detect the ubiquitin conjugating activity of the purified RAD6B protein on its substrate H2A. The results indicated that the RAD6B protein we purified was able to carry out ubiquitination reaction on H2A. This indicated that the purified protein was active. We were also able to obtain the crystals of RAD6B, which serves as a good starting point for structure determination. These crystals will further be optimized to grow bigger crystals, suitable for X-ray diffraction and data collection.

Further studies to solve the crystal structure of RAD6B/SMI9 complex will provide the excellent basis for a rational design of more potent and selective small molecule inhibitors. New RAD6B inhibitors with improved potency and selectivity can be designed by using a combined

ligand and receptor-based method. In summary, our study has paved the way for developing novel, small and selective molecule inhibitors of RAD6B which can be used to treat breast cancer in future.

REFERENCES

1. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.* Nat Genet, 2003. **33 Suppl**: p. 245-54.
2. *https://www2.le.ac.uk/projects/vgec/highereducation/epigenetics_ethics/Introduction*.
3. *https://www.zymoresearch.com/pages/intro-to-epigenetics*.
4. *https://edu.t-bio.info/course/epigenetics-1/*.
5. *https://ysjournal.com/role-of-epigenetic-modifications-on-mitotic-chromosome-structure/*.
6. Lee, J., G. F Murphy, and C. Lian, *Melanoma Epigenetics: Novel Mechanisms, Markers, and Medicines*. Vol. 94. 2014.
7. He, C. and P. Cole, *Introduction: epigenetics.* Chemical reviews, 2015. **115**(6): p. 2223-2224.
8. *https://courses.lumenlearning.com/wsu-sandbox/chapter/introduction-to-epigenetics/*.
9. Jin, B., Y. Li, and K.D. Robertson, *DNA methylation: superior or subordinate in the epigenetic hierarchy?* Genes Cancer, 2011. **2**(6): p. 607-17.
10. Jones, P.A. and D. Takai, *The Role of DNA Methylation in Mammalian Epigenetics.* Science, 2001. **293**(5532): p. 1068-1070.
11. Ballestar, E., *An introduction to epigenetics.* Adv Exp Med Biol, 2011. **711**: p. 1-11.
12. *https://www.whatisepigenetics.com/histone-modifications/*.
13. *https://www.abcam.com/epigenetics/histone-modifications-a-guide*.
14. *https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-post-translational-modification.html*.

15. *https://www.news-medical.net/life-sciences/Types-of-Protein-Post-Translational-Modification.aspx.*

16. *https://docs.abcam.com/pdf/chromatin/a-guide-to-epigenetics.pdf.*

17. Martin, C. and Y. Zhang, *The diverse functions of histone lysine methylation.* Nature Reviews Molecular Cell Biology, 2005. **6**: p. 838.

18. Nagalakshmi, B., S. Sagarkar, and A.J. Sakharkar, *Epigenetic Mechanisms of Traumatic Brain Injuries*, in *Progress in Molecular Biology and Translational Science*, D.R. Grayson, Editor. 2018, Academic Press. p. 263-298.

19. *https://www.news-medical.net/life-sciences/Ubiquitination-(Ubiquitylation).aspx.*

20. Stringer, D.K. and R.C. Piper, *Terminating protein ubiquitination: Hasta la vista, ubiquitin.* Cell cycle (Georgetown, Tex.), 2011. **10**(18): p. 3067-3071.

21. Pickart, C.M. and M.J. Eddins, *Ubiquitin: structures, functions, mechanisms.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2004. **1695**(1): p. 55-72.

22. Wu, Z., J. Connolly, and K.K. Biggar, *Beyond histones - the expanding roles of protein lysine methylation.* Febs j, 2017. **284**(17): p. 2732-2744.

23. Li, J., et al., *Role of several histone lysine methyltransferases in tumor development.* Biomedical reports, 2016. **4**(3): p. 293-299.

24. Liu, Q. and M.-w. Wang, *Histone lysine methyltransferases as anti-cancer targets for drug discovery.* Acta Pharmacologica Sinica, 2016. **37**: p. 1273.

25. Spellmon, N., et al., *Structure and function of SET and MYND domain-containing proteins.* International journal of molecular sciences, 2015. **16**(1): p. 1406-1428.

26. Du, S.J., X. Tan, and J. Zhang, *SMYD proteins: key regulators in skeletal and cardiac muscle development and function.* Anat Rec (Hoboken), 2014. **297**(9): p. 1650-62.

27. Sirinupong, N., et al., *Crystal structure of cardiac-specific histone methyltransferase SmyD1 reveals unusual active site architecture.* J Biol Chem, 2010. **285**(52): p. 40635-44.

28. Feng, Q., et al., *Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain.* Curr Biol, 2002. **12**(12): p. 1052-8.

29. Green, E.M., A.J. Morrison, and O. Gozani, *New marks on the block.* Nucleus, 2012. **3**(4): p. 335-339.

30. Boriack-Sjodin, P.A. and K.K. Swinger, *Protein Methyltransferases: A Distinct, Diverse, and Dynamic Family of Enzymes.* Biochemistry, 2016. **55**(11): p. 1557-69.

31. Hamey, J.J., et al., *Novel N-terminal and Lysine Methyltransferases That Target Translation Elongation Factor 1A in Yeast and Human.* Molecular & Cellular Proteomics, 2016. **15**(1): p. 164.

32. Dzialo, M.C., et al., *Translational roles of elongation factor 2 protein lysine methylation.* J Biol Chem, 2014. **289**(44): p. 30511-24.

33. Weiss, V.H., et al., *The structure and oligomerization of the yeast arginine methyltransferase, Hmt1.* Nat Struct Biol, 2000. **7**(12): p. 1165-71.

34. Ozkaynak, E., et al., *The yeast ubiquitin genes: a family of natural gene fusions.* Embo j, 1987. **6**(5): p. 1429-39.

35. Jentsch, S., J. McGrath, and A. Varshavsky, *The DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme*. Vol. 329. 1987. 131-4.

36. Koken, M.H.M., et al., *Localization of two human homologs, HHR6A and HHR6B, of the yeast DNA repair gene RAD6 to chromosomes Xq24–q25 and 5q23–q31*. Genomics, 1992. **12**(3): p. 447-453.

37. Komander, D. and M. Rape, *The Ubiquitin Code.* Annual Review of Biochemistry, 2012. **81**(1): p. 203-229.

38. Kumar, P., et al., *Role of a non-canonical surface of Rad6 in ubiquitin conjugating activity.* Nucleic Acids Res, 2015. **43**(18): p. 9039-50.

39. Watkins, J.F., et al., *The extremely conserved amino terminus of RAD6 ubiquitin-conjugating enzyme is essential for amino-end rule-dependent protein degradation.* Genes Dev, 1993. **7**(2): p. 250-61.

40. Shekhar, M.P., et al., *Rad6 overexpression induces multinucleation, centrosome amplification, abnormal mitosis, aneuploidy, and transformation.* Cancer Res, 2002. **62**(7): p. 2115-24.

41. Shekhar, M.P., L. Tait, and B. Gerard, *Essential role of T-cell factor/beta-catenin in regulation of Rad6B: a potential mechanism for Rad6B overexpression in breast cancer cells.* Mol Cancer Res, 2006. **4**(10): p. 729-45.

42. Howe, L.R. and A.M.C. Brown, *Wnt Signaling and Breast Cancer.* Cancer Biology & Therapy, 2004. **3**(1): p. 36-41.

43. Brown, A.M., *Wnt signaling in breast cancer: have we come full circle?* Breast Cancer Res, 2001. **3**(6): p. 351-5.

44. Shekhar, M.P., et al., *Rad6B is a positive regulator of beta-catenin stabilization.* Cancer Res, 2008. **68**(6): p. 1741-50.

45. Sharp, P.M. and W.H. Li, *Ubiquitin genes as a paradigm of concerted evolution of tandem repeats.* J Mol Evol, 1987. **25**(1): p. 58-64.

46. Mongeon, V., *Elucidating the biochemical and structural features required for SMYD5 mediated methylation of histone H4 and other potential substrates.* 2014.

47. Xue, W., *Thesis: Expression, Purification and Characterization of Lysine Methyltransferase SMYD5.* 2017.

48. Duchin, S., et al., *A continuous kinetic assay for protein and DNA methyltransferase enzymatic activities.* Epigenetics Chromatin, 2015. **8**: p. 56.

49. Green, E.M., et al., *Methylation of H4 lysines 5, 8 and 12 by yeast Set5 calibrates chromatin stress responses.* Nat Struct Mol Biol, 2012. **19**(3): p. 361-3.

50. Jezek, M., et al., *The histone methyltransferases Set5 and Set1 have overlapping functions in gene silencing and telomere maintenance.* Epigenetics, 2017. **12**(2): p. 93-104.

51. Martin, G.M., et al., *Set5 and Set1 cooperate to repress gene expression at telomeres and retrotransposons.* Epigenetics, 2014. **9**(4): p. 513-22.

52. Smyth, M.S. and J.H. Martin, *x ray crystallography.* Molecular pathology : MP, 2000. **53**(1): p. 8-14.

53. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2019.* CA Cancer J Clin, 2019. **69**(1): p. 7-34.

54. Torre, L.A., et al., *Global cancer statistics, 2012.* CA Cancer J Clin, 2015. **65**(2): p. 87-108.

55. Tomlinson, I.P., M.R. Novelli, and W.F. Bodmer, *The mutation rate and cancer.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(25): p. 14800-14803.

56. Paula Gill, A.G., and Charles Loprinzi, *Nausea and Vomiting in the Cancer Patient.*

57. Nounou, M.I., et al., *Breast cancer: conventional diagnosis and treatment modalities and recent patents and technologies.* Breast cancer: basic and clinical research, 2015. **9**: p. BCBCR. S29420.

58. Sanders, M.A., et al., *Novel inhibitors of Rad6 ubiquitin conjugating enzyme: design, synthesis, identification, and functional characterization.* Mol Cancer Ther, 2013. **12**(4): p. 373-83.

59. Sanders, M.A., et al., *Pharmacological targeting of RAD6 enzyme-mediated translesion synthesis overcomes resistance to platinum-based drugs.* J Biol Chem, 2017. **292**(25): p. 10347-10363.

60. Hibbert, R.G., et al., *E3 ligase Rad18 promotes monoubiquitination rather than ubiquitin chain formation by E2 enzyme Rad6.* Proc Natl Acad Sci U S A, 2011. **108**(14): p. 5590-5.

61. Aberle, H., et al., *beta-catenin is a target for the ubiquitin-proteasome pathway.* Embo j, 1997. **16**(13): p. 3797-804.

ABSTRACT

STRUCTURAL INSIGHTS INTO HISTONE MODIFYING ENZYMES

by

SHRUTI AMLE

MAY 2019

Advisor: Dr. Zhe Yang

Major: Biochemistry and Molecular Biology

Degree: Master of Science

X-ray crystallography is a powerful tool in the elucidation of the three-dimensional structure of bio molecules such as proteins and nucleic acids. Obtaining their structures will allow deep understanding of their biochemical mechanisms and how the alterations in the structures influence human health. X-crystallography has been used as an important approach to understand the structural and biochemical properties of histone modifying enzymes and shed light on poorly understood mechanisms.

SMYD5, one of the members of SMYD protein family, is identified as a histone lysine methyltransferase. Lysine methylation is known to modulate various biological processes including DNA damage response and gene regulation. SMYD5 is involved in hematopoiesis regulation, immune response and tumorigenesis. So far, the biochemical and structural features required to understand its methyl transferase function remains elusive. Solving the crystal structure and identifying the new substrates of SMYD5 will provide a molecular window to visualize how these biological molecules influence human health. In this study, we conducted the crystallization attempts and various biochemical assays to understand the function of SMYD5. To identify if SMYD5 can methylate non histone substrates, an artificial peptide, assumed to be the best peptide

substrate for SMYD5 was designed based on the substrate selectivity profile of SMYD5. We found that SMYD5 is much more active on this peptide than its known histone substrate H4K20 in both enzyme kinetics assay and dot blot assay.

Methylation controls chromatin regulation which is catalyzed by SET domain-containing methyltransferases. SET5 was identified as the first histone methyltransferase enzyme in budding yeast. It is known to monomethylate the important H4 lysine residues, H4K5, H4K8 and H4K12. Studies have been demonstrated a functional relationship between histone H3 and histone H4 methyltransferases, whose combined activities play roles in preserving genomic integrity. Also, SET5 is recognized as human SMYD3 ortholog, which also catalyzes the methylation at H4K5. Therefore, SET5 functional studies might reveal previously unknown conserved mechanisms that may contribute to SMYD3 dependent oncogenesis in human cells. In this study, X-ray crystallographic studies were utilized to solve the crystal structure of SET5. Our initial crystallization screening has successfully resulted in Set5 crystallization. Solving its crystal structure will shed light on poorly understood mechanisms of SET5-mediated histone H4 methylation.

Rad6B plays a key role in breast cancer cell growth. It is a ubiquitin conjugating enzyme (E2) which mediates ubiquitination. It is essential for the post-replication, DNA repair and genomic integrity maintenance via its ubiquitin-conjugating activity. Alterations in ubiquitination occur frequently in cancer. Because Rad6B is overexpressed in breast cancer cells, targeting Rad6B would be a viable approach for breast cancer treatment. SMI9 is a specific small molecule inhibitor of Rad6B. It is known to inhibit Rad6B-induced histone H2A ubiquitination, induce G2–M arrest and apoptosis, downregulate intracellular β-catenin and inhibit proliferation and migration of metastatic human breast cancer cells. In our study, the crystals were obtained for

Rad6B which will serve as a good starting point to and solve macromolecular structure of Rad6B in complex with SMI9. Solving the crystal structure will facilitate the development of new Rad6B inhibitors which can be used to inhibit the overexpression of Rad6B in breast cancer.

AUTOBIOGRAPHICAL STATEMENT

Education

2008-2011 B.Sc. Biotechnology, Marathwada University, India

2011-2013 M.Sc. Biotechnology, Marathwada University, India

2017-present M.Sc. Biochemistry and Molecular Biology; Wayne State University, Detroit, MI